Omics\_PEC2

Judith Guitart

1/6/2020

# Preparación de los datos

Primeramente se leen los archivos ‘targets.csv’ y ‘counts.csv’ que se encuentran en la carpeta ‘Datos’ dentro de la carpeta ‘PEC2’ de la asignatura, establecida como el directorio principal de este análisis.

targets <- read.csv("./Datos/targets.csv", header = TRUE, sep = ",", row.names=1)  
head(targets)

## SRA\_Sample Sample\_Name Grupo\_analisis body\_site  
## SRX567480 SRS626942 GTEX-111CU-0226-SM-5GZXC 1 Thyroid  
## SRX615964 SRS644174 GTEX-111FC-1026-SM-5GZX1 1 Thyroid  
## SRX563960 SRS625636 GTEX-111VG-0526-SM-5N9BW 3 Thyroid  
## SRX564185 SRS625665 GTEX-111YS-0726-SM-5GZY8 1 Thyroid  
## SRX559141 SRS624025 GTEX-1122O-0226-SM-5N9DA 1 Thyroid  
## SRX561718 SRS625313 GTEX-1128S-0126-SM-5H12S 1 Thyroid  
## molecular\_data\_type sex Group ShortName  
## SRX567480 Allele-Specific Expression male NIT 111CU\_NIT  
## SRX615964 RNA Seq (NGS) male NIT 111FC\_NIT  
## SRX563960 RNA Seq (NGS) male ELI 111VG\_ELI  
## SRX564185 Allele-Specific Expression male NIT 111YS\_NIT  
## SRX559141 RNA Seq (NGS) female NIT 1122O\_NIT  
## SRX561718 Allele-Specific Expression female NIT 1128S\_NIT

counts <- read.csv("./Datos/counts.csv", header = TRUE, sep = ";", row.names = 1)

A continuación, se extraen 10 muestras de cada uno de los grupos (NIT, SFI y ELI) de manera aleatoria del archivo ‘targets.csv’ mediante el paquete dplyr:

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

set.seed(3333)  
targets.pec <- targets %>% group\_by(Group) %>% sample\_n(10) %>% arrange(Grupo\_analisis)  
str(targets.pec)

## tibble [30 x 8] (S3: grouped\_df/tbl\_df/tbl/data.frame)  
## $ SRA\_Sample : chr [1:30] "SRS629372" "SRS629803" "SRS627199" "SRS389744" ...  
## $ Sample\_Name : chr [1:30] "GTEX-11ZTT-1026-SM-5EQKF" "GTEX-11GSO-0626-SM-5A5LW" "GTEX-Y5V5-0326-SM-5RQJG" "GTEX-QEL4-0726-SM-3GIJ5" ...  
## $ Grupo\_analisis : int [1:30] 1 1 1 1 1 1 1 1 1 1 ...  
## $ body\_site : chr [1:30] "Thyroid" "Thyroid" "Thyroid" "Thyroid" ...  
## $ molecular\_data\_type: chr [1:30] "Allele-Specific Expression" "RNA Seq (NGS)" "RNA Seq (NGS)" "Allele-Specific Expression" ...  
## $ sex : chr [1:30] "female" "male" "female" "male" ...  
## $ Group : chr [1:30] "NIT" "NIT" "NIT" "NIT" ...  
## $ ShortName : chr [1:30] "11ZTT\_NIT" "11GSO\_NIT" "Y5V5-\_NIT" "QEL4-\_NIT" ...  
## - attr(\*, "groups")= tibble [3 x 2] (S3: tbl\_df/tbl/data.frame)  
## ..$ Group: chr [1:3] "ELI" "NIT" "SFI"  
## ..$ .rows: list<int> [1:3]   
## .. ..$ : int [1:10] 21 22 23 24 25 26 27 28 29 30  
## .. ..$ : int [1:10] 1 2 3 4 5 6 7 8 9 10  
## .. ..$ : int [1:10] 11 12 13 14 15 16 17 18 19 20  
## .. ..@ ptype: int(0)   
## ..- attr(\*, ".drop")= logi TRUE

Se observa que el output es el deseado.

A continuación, se quiere subsetear estas muestras escogidas en las columnas del archivo ‘counts.csv’ para obtener la información adecuada de cada una de las muestras.

Para ello, es necesario que los nombres de la columna ‘Sample\_Name’ coincidan con los nombres de cada una de las columnas del archivo ‘counts.csv’. Para ello, es posible utilizar la función gsub para sustituir el carácter . por el carácter -.

names(counts) <- gsub(x=names(counts), pattern = "\\.", replacement = "-")

Para obtener el archivo ‘counts.pec’ con los datos de ‘counts’ de las muestras seleccionadas en el archivo ‘targets.pec’, se usa la función select del paquete dyplr:

counts.pec <- counts %>% select(one\_of(as.character(targets.pec$Sample\_Name)))  
str(counts.pec)

## 'data.frame': 56202 obs. of 30 variables:  
## $ GTEX-11ZTT-1026-SM-5EQKF: int 3 580 0 1 1 0 2 3 13 879 ...  
## $ GTEX-11GSO-0626-SM-5A5LW: int 0 544 0 3 0 0 2 4 5 1129 ...  
## $ GTEX-Y5V5-0326-SM-5RQJG : int 1 424 0 2 1 0 0 10 5 241 ...  
## $ GTEX-QEL4-0726-SM-3GIJ5 : int 4 511 2 2 4 1 0 5 50 477 ...  
## $ GTEX-RU1J-0226-SM-2TF5Y : int 4 415 1 3 1 2 1 1 161 262 ...  
## $ GTEX-111CU-0226-SM-5GZXC: int 7 401 4 2 0 0 0 6 16 744 ...  
## $ GTEX-XUW1-1026-SM-4BONY : int 2 688 1 0 0 0 0 6 17 868 ...  
## $ GTEX-X4XY-0826-SM-4E3JM : int 1 1071 1 3 0 0 4 1 23 619 ...  
## $ GTEX-133LE-0326-SM-5P9G4: int 4 1175 2 3 2 0 4 4 37 2739 ...  
## $ GTEX-Y111-1926-SM-4SOIS : int 2 810 0 2 0 1 1 6 10 993 ...  
## $ GTEX-13NZ8-0226-SM-5J2OK: int 1 1164 2 2 2 0 0 19 5 181 ...  
## $ GTEX-117YW-0126-SM-5EGGN: int 1 483 0 0 0 0 0 0 10 171 ...  
## $ GTEX-13FH7-0126-SM-5KLZ1: int 5 576 4 3 0 1 2 4 26 772 ...  
## $ GTEX-Q2AH-0726-SM-2I3EA : int 1 874 8 2 0 1 3 4 31 1086 ...  
## $ GTEX-139UW-0126-SM-5KM1B: int 2 430 0 0 0 0 0 9 9 679 ...  
## $ GTEX-131XG-0226-SM-5IFG1: int 0 325 1 2 0 2 1 13 21 673 ...  
## $ GTEX-11TUW-0226-SM-5LU8X: int 4 627 0 1 0 0 1 6 18 740 ...  
## $ GTEX-TKQ1-0126-SM-33HB3 : int 8 1388 1 1 1 0 3 8 40 410 ...  
## $ GTEX-12ZZY-0826-SM-5EQMT: int 2 491 2 0 0 0 1 7 4 552 ...  
## $ GTEX-13FXS-0726-SM-5LZXJ: int 5 1564 0 2 1 0 1 3 8 283 ...  
## $ GTEX-TMMY-0826-SM-33HB9 : int 3 979 3 2 5 8 3 4 81 527 ...  
## $ GTEX-13QJC-0826-SM-5RQKC: int 0 825 1 0 0 1 1 10 21 853 ...  
## $ GTEX-ZYY3-1926-SM-5GZXS : int 6 1003 1 2 0 1 4 8 12 960 ...  
## $ GTEX-YFC4-2626-SM-5P9FQ : int 1 1472 1 0 0 1 2 38 24 2020 ...  
## $ GTEX-11XUK-0226-SM-5EQLW: int 0 419 0 1 0 0 0 1 32 1002 ...  
## $ GTEX-R55G-0726-SM-2TC6J : int 3 134 1 2 1 0 1 3 143 366 ...  
## $ GTEX-PLZ4-1226-SM-2I5FE : int 5 489 1 3 2 1 3 7 100 523 ...  
## $ GTEX-14ABY-0926-SM-5Q5DY: int 1 775 2 0 0 0 1 10 2 580 ...  
## $ GTEX-YJ89-0726-SM-5P9F7 : int 4 1325 1 0 2 1 2 4 8 853 ...  
## $ GTEX-13NZ9-1126-SM-5MR37: int 0 1002 1 0 0 1 0 15 19 602 ...

Esta matriz de contaje contiene en cada fila un gen con el código Ensembl y cada una de las librerias RNA secuenciadas en las columnas con los respectivos valores asignados.

Es posible comprobar que las filas del archivo ‘targets.pec’ coincida con las columnas del archivo ‘counts.pec’:

all(rownames(targets.pec$Sample\_Name) == colnames(counts.pec))

## [1] TRUE

write.csv(targets.pec, "./Datos/targets.pec.csv")  
write.csv(counts.pec, "./Datos/counts.pec.csv")

A continuación, se crea la matriz del objeto DESeqDataSet que consta de los contajes (counts.pec) y la tabla de información de las muestas (targets.pec), mediante funciones del paquete *DESeq2*. La fórmula de diseño define qué columnas de la tabla de información de las muestras (targets.pec) especifica el diseño experimental y cómo estos factores deben usarse en el análisis. Como queremos determinar el efecto de los distintos grupos NIT, SFI y ELI, la fórmula de diseño será ~ Group.

if (!requireNamespace("BiocManager", quietly = TRUE))  
 install.packages("BiocManager")  
BiocManager::install("DESeq2")

## Bioconductor version 3.11 (BiocManager 1.30.10), R 4.0.0 (2020-04-24)

## Installing package(s) 'DESeq2'

## package 'DESeq2' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'DESeq2'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problema al copiar C:  
## \Users\Judith\Documents\R\win-library\4.0\00LOCK\DESeq2\libs\i386\DESeq2.dll  
## a C:\Users\Judith\Documents\R\win-library\4.0\DESeq2\libs\i386\DESeq2.dll:  
## Permission denied

## Warning: restored 'DESeq2'

##   
## The downloaded binary packages are in  
## C:\Users\Judith\AppData\Local\Temp\RtmpKizBus\downloaded\_packages

## Installation path not writeable, unable to update packages: boot, class,  
## foreign, KernSmooth, MASS, nlme, nnet, spatial, survival

## Old packages: 'limma', 'vctrs'

library("DESeq2")

## Loading required package: S4Vectors

## Loading required package: stats4

## Loading required package: BiocGenerics

## Loading required package: parallel

##   
## Attaching package: 'BiocGenerics'

## The following objects are masked from 'package:parallel':  
##   
## clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,  
## clusterExport, clusterMap, parApply, parCapply, parLapply,  
## parLapplyLB, parRapply, parSapply, parSapplyLB

## The following objects are masked from 'package:dplyr':  
##   
## combine, intersect, setdiff, union

## The following objects are masked from 'package:stats':  
##   
## IQR, mad, sd, var, xtabs

## The following objects are masked from 'package:base':  
##   
## anyDuplicated, append, as.data.frame, basename, cbind, colnames,  
## dirname, do.call, duplicated, eval, evalq, Filter, Find, get, grep,  
## grepl, intersect, is.unsorted, lapply, Map, mapply, match, mget,  
## order, paste, pmax, pmax.int, pmin, pmin.int, Position, rank,  
## rbind, Reduce, rownames, sapply, setdiff, sort, table, tapply,  
## union, unique, unsplit, which, which.max, which.min

##   
## Attaching package: 'S4Vectors'

## The following objects are masked from 'package:dplyr':  
##   
## first, rename

## The following object is masked from 'package:base':  
##   
## expand.grid

## Loading required package: IRanges

##   
## Attaching package: 'IRanges'

## The following objects are masked from 'package:dplyr':  
##   
## collapse, desc, slice

## The following object is masked from 'package:grDevices':  
##   
## windows

## Loading required package: GenomicRanges

## Loading required package: GenomeInfoDb

## Loading required package: SummarizedExperiment

## Loading required package: Biobase

## Welcome to Bioconductor  
##   
## Vignettes contain introductory material; view with  
## 'browseVignettes()'. To cite Bioconductor, see  
## 'citation("Biobase")', and for packages 'citation("pkgname")'.

## Loading required package: DelayedArray

## Loading required package: matrixStats

##   
## Attaching package: 'matrixStats'

## The following objects are masked from 'package:Biobase':  
##   
## anyMissing, rowMedians

## The following object is masked from 'package:dplyr':  
##   
## count

##   
## Attaching package: 'DelayedArray'

## The following objects are masked from 'package:matrixStats':  
##   
## colMaxs, colMins, colRanges, rowMaxs, rowMins, rowRanges

## The following objects are masked from 'package:base':  
##   
## aperm, apply, rowsum

DESeq2 offers transformations for count data that stabilize the variance across the mean: the regularized logarithm (rlog) and the variance stabilizing transformation (VST). These have slightly different implementations, discussed a bit in the DESeq2 paper and in the very extensive web tutorial, but a similar goal of stablizing the variance across the range of values. Both produce log2-like values for high counts. Here we will use the regularized log transformation implemented with the rlog function.

ddsM <- DESeqDataSetFromMatrix(countData = counts.pec, colData = targets.pec, design = ~Group)

## Warning in DESeqDataSet(se, design = design, ignoreRank): some variables in  
## design formula are characters, converting to factors

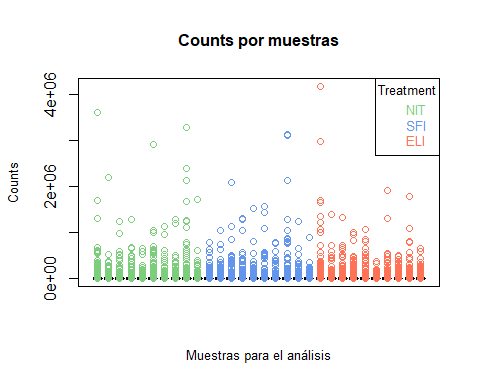
ddsM

## class: DESeqDataSet   
## dim: 56202 30   
## metadata(1): version  
## assays(1): counts  
## rownames(56202): ENSG00000223972.4 ENSG00000227232.4 ...  
## ENSG00000210195.2 ENSG00000210196.2  
## rowData names(0):  
## colnames(30): GTEX-11ZTT-1026-SM-5EQKF GTEX-11GSO-0626-SM-5A5LW ...  
## GTEX-YJ89-0726-SM-5P9F7 GTEX-13NZ9-1126-SM-5MR37  
## colData names(8): SRA\_Sample Sample\_Name ... Group ShortName

# Preprocesado de los datos: filtraje y normalización

Con el objetivo de visualizar relaciones entre las muestras, primeramente se van a realizar transformaciones en los contajes y seguidamente se van a realizar test estadísticos sobre los mismos.

boxplot(counts.pec, names = c(rep(" ",30)), outcol = c(rep("palegreen3",10),rep("cornflowerblue",10),rep("coral1",10)), xaxt="n", main = "Counts por muestras", xlab = "Muestras para el análisis", ylab = "Counts", cex.lab = 0.8, cex.main=1)  
legend("topright",legend=c("NIT","SFI","ELI"), text.col=c("palegreen3","cornflowerblue","coral1"), cex=0.8, title = "Treatment", title.col = "black")



## Pre-filtrado de los datos

Este paso consiste en eliminar esas líneas que menos de 10 reads con el objetivo de reducir las dimensiones del objeto de datos y así incrementar la velocidad de las funciones, ya que no aportan información al análisis. (que seria un cpm de 0.5)

Así, se observa que el objeto inicial contiene 56202 líneas.

nrow(ddsM)

## [1] 56202

Para filtrar estos datos, se seleccionan todas aquellas líneas que contienen almenos dos contajes en las treinta muestras.

dds <- ddsM[rowSums(counts(ddsM)) >= 10, ]  
nrow(dds)

## [1] 35894

Ahora se observa que el objeto contiene 36091 líneas, es decir, se han eliminado 12815 códigos de genes que no aportan información de expresión diferencial al análisis.

For transformation: The point of these two transformations, the VST and the rlog, is to remove the dependence of the variance on the mean, particularly the high variance of the logarithm of count data when the mean is low. Both VST and rlog use the experiment-wide trend of variance over mean, in order to transform the data to remove the experiment-wide trend. Note that we do not require or desire that all the genes have exactly the same variance after transformation. Indeed, in a figure below, you will see that after the transformations the genes with the same mean do not have exactly the same standard deviations, but that the experiment-wide trend has flattened. It is those genes with row variance above the trend which will allow us to cluster samples into interesting groups.

## Transformación para varianza estable y rlog

Muchos métodos estadísticos requieren que la varianza sea homogénea entre las muestras, es decir, requieren homocedasticidad. El paquete *DESeq2* ofrece dos transformaciones para contajes de RNA-seq para estabilizar la varianza: la función *Vst* y la función *rlog*.

Se realiza primeramente una transformación con la función *vst*, ya que nuestros datos contienen 30 muestras, con el atributo blind en *FALSE*, pues en esta primera transformación no quieren establecerse diferencias entre los grupos.

vsd <- vst(dds, blind=FALSE)

A ctontinuación se realiza la otra función de transformación:

rld <- rlog(dds, blind=FALSE)

## rlog() may take a few minutes with 30 or more samples,  
## vst() is a much faster transformation

Para mostrar el efecto de la transformación se muestran en los siguientes plots:

BiocManager::install("hexbin")

## Bioconductor version 3.11 (BiocManager 1.30.10), R 4.0.0 (2020-04-24)

## Installing package(s) 'hexbin'

## package 'hexbin' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'hexbin'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problema al copiar C:  
## \Users\Judith\Documents\R\win-library\4.0\00LOCK\hexbin\libs\i386\hexbin.dll  
## a C:\Users\Judith\Documents\R\win-library\4.0\hexbin\libs\i386\hexbin.dll:  
## Permission denied

## Warning: restored 'hexbin'

##   
## The downloaded binary packages are in  
## C:\Users\Judith\AppData\Local\Temp\RtmpKizBus\downloaded\_packages

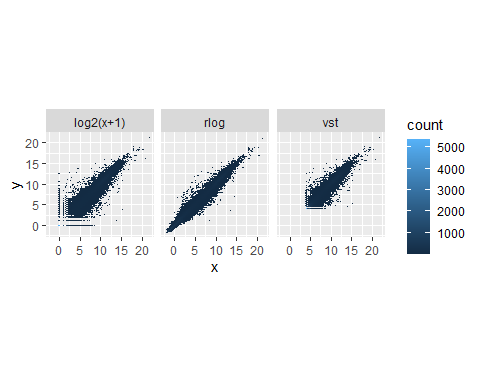
## Installation path not writeable, unable to update packages: boot, class,  
## foreign, KernSmooth, MASS, nlme, nnet, spatial, survival

## Old packages: 'limma', 'vctrs'

library("ggplot2")  
library("hexbin")  
ddsN <- estimateSizeFactors(dds)  
df <- bind\_rows(  
 as\_data\_frame(log2(counts(ddsN, normalized=TRUE)[, 1:2]+1)) %>% mutate(transformation = "log2(x+1)"),  
 as\_data\_frame(assay(vsd)[, 1:2]) %>% mutate(transformation = "vst"), as\_data\_frame(assay(rld)[, 1:2]) %>% mutate(transformation = "rlog")  
)

## Warning: `as\_data\_frame()` is deprecated as of tibble 2.0.0.  
## Please use `as\_tibble()` instead.  
## The signature and semantics have changed, see `?as\_tibble`.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_warnings()` to see where this warning was generated.

colnames(df)[1:2] <- c("x","y")  
ggplot(df, aes(x = x, y = y)) + geom\_hex(bins = 80) + coord\_fixed() + facet\_grid( . ~transformation)



A la izquierda se observan los contajes normalizados transformados log2, con la transformación rlog en el medio y a la derecha se encuentran los datos transformados con VST. Se observa como las transformaciones vst y rlog muestran diferencias por los genes con muy bajos contajes, ya que estos datos dan poca información sobre expresión diferencial.

The trend typically captures high dispersions for low counts, and therefore these genes exhibit higher shrinkage from the rlog.

## Normalization:

BiocManager::install("edgeR")

## Bioconductor version 3.11 (BiocManager 1.30.10), R 4.0.0 (2020-04-24)

## Installing package(s) 'edgeR'

## package 'edgeR' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'edgeR'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problema al copiar C:  
## \Users\Judith\Documents\R\win-library\4.0\00LOCK\edgeR\libs\i386\edgeR.dll a C:  
## \Users\Judith\Documents\R\win-library\4.0\edgeR\libs\i386\edgeR.dll: Permission  
## denied

## Warning: restored 'edgeR'

##   
## The downloaded binary packages are in  
## C:\Users\Judith\AppData\Local\Temp\RtmpKizBus\downloaded\_packages

## Installation path not writeable, unable to update packages: boot, class,  
## foreign, KernSmooth, MASS, nlme, nnet, spatial, survival

## Old packages: 'limma', 'vctrs'

library("edgeR")

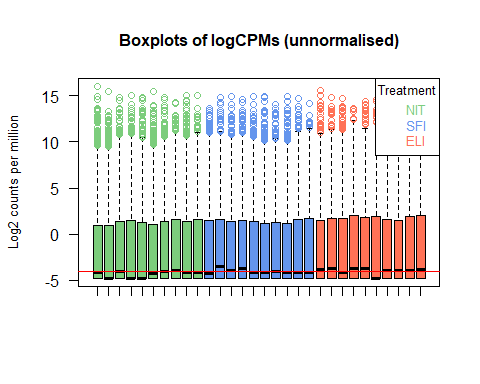
## Loading required package: limma

##   
## Attaching package: 'limma'

## The following object is masked from 'package:DESeq2':  
##   
## plotMA

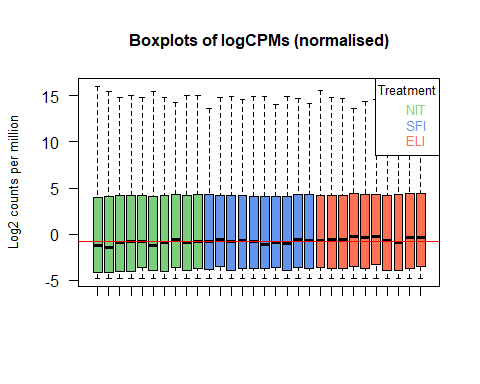
## The following object is masked from 'package:BiocGenerics':  
##   
## plotMA

logcounts <- cpm(counts.pec, log=TRUE)  
boxplot(logcounts, names = c(rep(" ",30)), col = c(rep("palegreen3",10),rep("cornflowerblue",10),rep("coral1",10)), outcol = c(rep("palegreen3",10),rep("cornflowerblue",10),rep("coral1",10)), ylab = "Log2 counts per million", las=2, cex.lab = 0.8)  
abline(h=median(logcounts), col="red")  
title("Boxplots of logCPMs (unnormalised)", cex.main = 1)  
legend("topright",legend=c("NIT","SFI","ELI"), text.col=c("palegreen3","cornflowerblue","coral1"), cex=0.8, title = "Treatment", title.col = "black")



ddsN = estimateSizeFactors(dds)

normcounts <- cpm(ddsN, log = TRUE)  
boxplot(normcounts, names = c(rep(" ",30)), col = c(rep("palegreen3",10),rep("cornflowerblue",10),rep("coral1",10)), ylab = "Log2 counts per million", las=2, cex.lab = 0.8)  
abline(h=median(normcounts), col="red")  
title("Boxplots of logCPMs (normalised)", cex.main = 1)  
legend("topright",legend=c("NIT","SFI","ELI"), text.col=c("palegreen3","cornflowerblue","coral1"), cex=0.8, title = "Treatment", title.col = "black")



Els valors de mediana son mas parecidos ahora.

# Visualización de los datos

## Distancias entre las muestras

sampleDists <- dist(t(assay(vsd)))

Ahora visualizamos estas distancias en un mapa de calor con la función pheatmap añadiendo las distancias entre las muestras como argumento para generar clusters entre las muestras.

BiocManager::install("pheatmap")

## Bioconductor version 3.11 (BiocManager 1.30.10), R 4.0.0 (2020-04-24)

## Installing package(s) 'pheatmap'

## package 'pheatmap' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\Judith\AppData\Local\Temp\RtmpKizBus\downloaded\_packages

## Installation path not writeable, unable to update packages: boot, class,  
## foreign, KernSmooth, MASS, nlme, nnet, spatial, survival

## Old packages: 'limma', 'vctrs'

BiocManager::install("RColorBrewer")

## Bioconductor version 3.11 (BiocManager 1.30.10), R 4.0.0 (2020-04-24)

## Installing package(s) 'RColorBrewer'

## package 'RColorBrewer' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\Judith\AppData\Local\Temp\RtmpKizBus\downloaded\_packages

## Installation path not writeable, unable to update packages: boot, class,  
## foreign, KernSmooth, MASS, nlme, nnet, spatial, survival

## Old packages: 'limma', 'vctrs'

library("pheatmap")  
library("RColorBrewer")  
sampleDistMatrix <- as.matrix(sampleDists)  
rownames(sampleDistMatrix) <- paste(vsd$Group, sep = " - ")  
colnames(sampleDistMatrix) <- NULL  
colors <- colorRampPalette(rev(brewer.pal(30, "Blues")) )(255)

## Warning in brewer.pal(30, "Blues"): n too large, allowed maximum for palette Blues is 9  
## Returning the palette you asked for with that many colors

pdf("pheatmap1.pdf")  
pheatmap1 <- pheatmap(sampleDistMatrix, clustering\_distance\_rows = sampleDists, clustering\_distance\_cols = sampleDists, col = colors)  
dev.off()

## pdf   
## 3

BiocManager::install("PoiClaClu")

## Bioconductor version 3.11 (BiocManager 1.30.10), R 4.0.0 (2020-04-24)

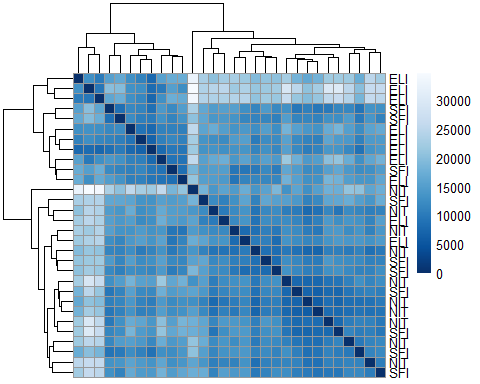
## Installing package(s) 'PoiClaClu'

## package 'PoiClaClu' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\Judith\AppData\Local\Temp\RtmpKizBus\downloaded\_packages

## Installation path not writeable, unable to update packages: boot, class,  
## foreign, KernSmooth, MASS, nlme, nnet, spatial, survival

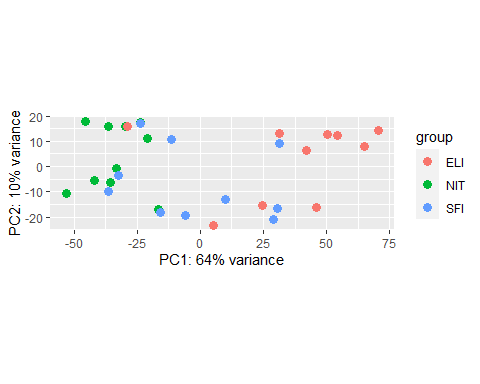
## Old packages: 'limma', 'vctrs'

library("PoiClaClu")  
poisd <- PoissonDistance(t(counts(dds)))  
samplePoisDistMatrix <- as.matrix( poisd$dd )  
rownames(samplePoisDistMatrix) <- paste( dds$Group, sep=" - " )  
colnames(samplePoisDistMatrix) <- NULL  
pheatmap(samplePoisDistMatrix,  
 clustering\_distance\_rows = poisd$dd,  
 clustering\_distance\_cols = poisd$dd,  
 col = colors)



## PCA Plot

plotPCA(vsd, intgroup = c("Group"))

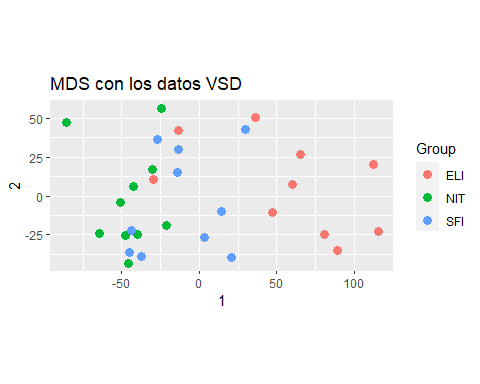


Se observa una separación entre los distintos grupos, aunque también hay algunas muestras que se disipan con otros grupos.

(afegir glm-pca?)

## MDS plot (Multidimensional scaling plots)

mds <- as.data.frame(colData(vsd)) %>% cbind(cmdscale(sampleDistMatrix))  
ggplot(mds, aes(x = `1`, y = `2`, color = Group)) + geom\_point(size=3) + coord\_fixed() + ggtitle("MDS con los datos VSD")



pdf("PCAandMDS.pdf")  
par(mfrow=c(1,2))  
plotPCA(vsd, intgroup = c("Group"))  
mds <- as.data.frame(colData(vsd)) %>% cbind(cmdscale(sampleDistMatrix))  
ggplot(mds, aes(x = `1`, y = `2`, color = Group)) + geom\_point(size=3) + coord\_fixed() + ggtitle("MDS con los datos VSD")  
dev.off()

## png   
## 2

# Análisis de expresión diferencial entre grupos

dds <- DESeq(dds, parallel = TRUE)

## estimating size factors

## estimating dispersions

## gene-wise dispersion estimates: 2 workers

## mean-dispersion relationship

## final dispersion estimates, fitting model and testing: 2 workers

## -- replacing outliers and refitting for 282 genes  
## -- DESeq argument 'minReplicatesForReplace' = 7   
## -- original counts are preserved in counts(dds)

## estimating dispersions

## fitting model and testing

La función DESeq lleva a cabo las siguientes tres funciones en los datos crudos: estimateSizeFactors(dds), estimateDispersions(dds) y nbinomWaldTest(dds). The results function of the DESeq2 package performs independent filtering by default using the mean of normalized counts as a filter statistic. A threshold on the filter statistic is found which optimizes the number of adjusted p values lower than a [specified] significance level”.

La variable *Group* contiene tres niveles, que pueden compararse dos a dos, mostrando así las tablas resultado de las distintas comparaciones: con un filtraje independiente de alpha 0.5.

## NIT vs SFI

res\_NITvsSFI <- results(dds, contrast=c("Group","NIT","SFI"), alpha=0.05)  
head(res\_NITvsSFI)

## log2 fold change (MLE): Group NIT vs SFI   
## Wald test p-value: Group NIT vs SFI   
## DataFrame with 6 rows and 6 columns  
## baseMean log2FoldChange lfcSE stat pvalue  
## <numeric> <numeric> <numeric> <numeric> <numeric>  
## ENSG00000223972.4 2.841267 0.217934 0.659522 0.330442 0.741066  
## ENSG00000227232.4 711.933057 -0.110665 0.235843 -0.469231 0.638905  
## ENSG00000243485.2 1.333301 -0.595119 0.816481 -0.728883 0.466073  
## ENSG00000237613.2 1.653664 0.860066 0.713099 1.206097 0.227780  
## ENSG00000268020.2 0.786348 1.362206 1.281722 1.062794 0.287876  
## ENSG00000240361.1 0.706012 0.168742 1.267970 0.133080 0.894130  
## padj  
## <numeric>  
## ENSG00000223972.4 0.963096  
## ENSG00000227232.4 0.947186  
## ENSG00000243485.2 NA  
## ENSG00000237613.2 NA  
## ENSG00000268020.2 NA  
## ENSG00000240361.1 NA

summary(res\_NITvsSFI)

##   
## out of 35890 with nonzero total read count  
## adjusted p-value < 0.05  
## LFC > 0 (up) : 16, 0.045%  
## LFC < 0 (down) : 285, 0.79%  
## outliers [1] : 0, 0%  
## low counts [2] : 7658, 21%  
## (mean count < 2)  
## [1] see 'cooksCutoff' argument of ?results  
## [2] see 'independentFiltering' argument of ?results

Se observan 112 genes sobreexpresados y 98 genes downregulados. Decir que se podria tambien añadir un lindar para el foldchange, con lfcThreshold=0.5, por ejemplo.

Si se considera una fracción de falsos positivos del 10% como aceptable se pueden ordenar mediatne el parámetro de fold change para obtener los genes significates con mayor downregulación y upregulación:

resSig\_NITvsSFI <- subset(res\_NITvsSFI, padj <= 0.05)  
head(resSig\_NITvsSFI[ order(resSig\_NITvsSFI$log2FoldChange), ])

## log2 fold change (MLE): Group NIT vs SFI   
## Wald test p-value: Group NIT vs SFI   
## DataFrame with 6 rows and 6 columns  
## baseMean log2FoldChange lfcSE stat pvalue  
## <numeric> <numeric> <numeric> <numeric> <numeric>  
## ENSG00000211930.1 18.1191 -8.32028 1.59002 -5.23283 1.66937e-07  
## ENSG00000211670.2 228.1669 -8.24635 1.14141 -7.22468 5.02276e-13  
## ENSG00000223648.2 80.7691 -7.91951 1.27513 -6.21076 5.27287e-10  
## ENSG00000211951.2 224.7878 -7.90541 1.19675 -6.60571 3.95627e-11  
## ENSG00000254395.1 73.6506 -7.86433 1.12051 -7.01852 2.24223e-12  
## ENSG00000224373.2 2417.2338 -7.84254 1.14513 -6.84861 7.45689e-12  
## padj  
## <numeric>  
## ENSG00000211930.1 5.17981e-05  
## ENSG00000211670.2 1.77278e-09  
## ENSG00000223648.2 3.16776e-07  
## ENSG00000211951.2 3.86031e-08  
## ENSG00000254395.1 5.75560e-09  
## ENSG00000224373.2 1.23855e-08

head(resSig\_NITvsSFI[ order(resSig\_NITvsSFI$log2FoldChange, decreasing = TRUE), ])

## log2 fold change (MLE): Group NIT vs SFI   
## Wald test p-value: Group NIT vs SFI   
## DataFrame with 6 rows and 6 columns  
## baseMean log2FoldChange lfcSE stat pvalue  
## <numeric> <numeric> <numeric> <numeric> <numeric>  
## ENSG00000110680.8 162.54852 4.32314 1.010264 4.27922 1.87548e-05  
## ENSG00000163501.6 139.22269 3.77893 0.838182 4.50848 6.52926e-06  
## ENSG00000240707.2 3.86898 2.90660 0.788996 3.68393 2.29669e-04  
## ENSG00000267131.1 9.88376 2.81455 0.549238 5.12446 2.98394e-07  
## ENSG00000247311.2 36.96961 2.38123 0.685129 3.47560 5.09714e-04  
## ENSG00000224965.1 42.93248 2.21548 0.620896 3.56820 3.59445e-04  
## padj  
## <numeric>  
## ENSG00000110680.8 0.003209461  
## ENSG00000163501.6 0.001280280  
## ENSG00000240707.2 0.026043874  
## ENSG00000267131.1 0.000087765  
## ENSG00000247311.2 0.048622607  
## ENSG00000224965.1 0.037589909

library("xlsx")  
write.csv(resSig\_NITvsSFI, "./Resultados/resSig\_NITvsSFI.csv")  
write.xlsx(resSig\_NITvsSFI, "./Resultados/resSig\_NITvsSFI.xlsx")

The first column, baseMean, is a just the average of the normalized count values, dividing by size factors, taken over all samples. The remaining four columns refer to a specific contrast, namely the comparison of the levels DPN versus Control of the factor variable treatment.

See the help page for results (by typing ?results) for information on how to obtain other contrasts.

The column log2FoldChange is the effect size estimate. It tells us how much the gene’s expression seems to have changed due to treatment with DPN in comparison to control.

NA = means that all counts for that gene are zero, and hence no test was applied.

Samllest p-value:

idx <- which.min(res\_NITvsSFI$pvalue)  
counts(dds, normalized = TRUE)[idx,]

## GTEX-11ZTT-1026-SM-5EQKF GTEX-11GSO-0626-SM-5A5LW GTEX-Y5V5-0326-SM-5RQJG   
## 0.000000 4.411896 13.278581   
## GTEX-QEL4-0726-SM-3GIJ5 GTEX-RU1J-0226-SM-2TF5Y GTEX-111CU-0226-SM-5GZXC   
## 0.000000 18.787488 4.453725   
## GTEX-XUW1-1026-SM-4BONY GTEX-X4XY-0826-SM-4E3JM GTEX-133LE-0326-SM-5P9G4   
## 10.857114 1.484383 12.618507   
## GTEX-Y111-1926-SM-4SOIS GTEX-13NZ8-0226-SM-5J2OK GTEX-117YW-0126-SM-5EGGN   
## 41.774814 123.696373 96.629375   
## GTEX-13FH7-0126-SM-5KLZ1 GTEX-Q2AH-0726-SM-2I3EA GTEX-139UW-0126-SM-5KM1B   
## 1200.249307 2028.684732 5895.806420   
## GTEX-131XG-0226-SM-5IFG1 GTEX-11TUW-0226-SM-5LU8X GTEX-TKQ1-0126-SM-33HB3   
## 80.367347 14.219545 200.234287   
## GTEX-12ZZY-0826-SM-5EQMT GTEX-13FXS-0726-SM-5LZXJ GTEX-TMMY-0826-SM-33HB9   
## 235.895462 1079.552450 2439.950733   
## GTEX-13QJC-0826-SM-5RQKC GTEX-ZYY3-1926-SM-5GZXS GTEX-YFC4-2626-SM-5P9FQ   
## 813.871711 2197.297258 3296.855747   
## GTEX-11XUK-0226-SM-5EQLW GTEX-R55G-0726-SM-2TC6J GTEX-PLZ4-1226-SM-2I5FE   
## 1804.397553 1959.534249 31.281933   
## GTEX-14ABY-0926-SM-5Q5DY GTEX-YJ89-0726-SM-5P9F7 GTEX-13NZ9-1126-SM-5MR37   
## 346.690970 636.687868 8581.356187

## NIT vs ELI

res\_NITvsELI <- results(dds, contrast=c("Group","NIT","ELI"), alpha = 0.05)  
head(res\_NITvsELI)

## log2 fold change (MLE): Group NIT vs ELI   
## Wald test p-value: Group NIT vs ELI   
## DataFrame with 6 rows and 6 columns  
## baseMean log2FoldChange lfcSE stat pvalue  
## <numeric> <numeric> <numeric> <numeric> <numeric>  
## ENSG00000223972.4 2.841267 0.50302620 0.667534 0.753559 0.4511143  
## ENSG00000227232.4 711.933057 0.00935783 0.235921 0.039665 0.9683602  
## ENSG00000243485.2 1.333301 0.11587157 0.848794 0.136513 0.8914156  
## ENSG00000237613.2 1.653664 1.27912774 0.731939 1.747588 0.0805354  
## ENSG00000268020.2 0.786348 0.29647296 1.207803 0.245465 0.8060967  
## ENSG00000240361.1 0.706012 -1.25282026 1.173849 -1.067276 0.2858473  
## padj  
## <numeric>  
## ENSG00000223972.4 0.688095  
## ENSG00000227232.4 0.986387  
## ENSG00000243485.2 0.954724  
## ENSG00000237613.2 0.246874  
## ENSG00000268020.2 NA  
## ENSG00000240361.1 NA

summary(res\_NITvsELI)

##   
## out of 35890 with nonzero total read count  
## adjusted p-value < 0.05  
## LFC > 0 (up) : 1525, 4.2%  
## LFC < 0 (down) : 3405, 9.5%  
## outliers [1] : 0, 0%  
## low counts [2] : 5571, 16%  
## (mean count < 1)  
## [1] see 'cooksCutoff' argument of ?results  
## [2] see 'independentFiltering' argument of ?results

Se observan 112 genes sobreexpresados y 98 genes downregulados.

Si se considera una fracción de falsos positivos del 10% como aceptable se pueden ordenar mediatne el parámetro de fold change para obtener los genes significates con mayor downregulación y upregulación:

resSig\_NITvsELI <- subset(res\_NITvsELI, padj <= 0.05)  
head(resSig\_NITvsELI[ order(resSig\_NITvsELI$log2FoldChange), ])

## log2 fold change (MLE): Group NIT vs ELI   
## Wald test p-value: Group NIT vs ELI   
## DataFrame with 6 rows and 6 columns  
## baseMean log2FoldChange lfcSE stat pvalue  
## <numeric> <numeric> <numeric> <numeric> <numeric>  
## ENSG00000211650.2 179.4440 -10.33621 1.39169 -7.42709 1.11010e-13  
## ENSG00000222037.5 898.6659 -9.49114 1.00183 -9.47382 2.69801e-21  
## ENSG00000211640.3 1252.9245 -9.32871 1.18764 -7.85485 4.00248e-15  
## ENSG00000211649.2 219.9166 -9.15017 1.21229 -7.54786 4.42478e-14  
## ENSG00000211951.2 224.7878 -8.98050 1.19655 -7.50533 6.12721e-14  
## ENSG00000234184.1 36.4657 -8.95778 1.40241 -6.38741 1.68715e-10  
## padj  
## <numeric>  
## ENSG00000211650.2 1.59533e-11  
## ENSG00000222037.5 1.02265e-17  
## ENSG00000211640.3 9.55647e-13  
## ENSG00000211649.2 6.95195e-12  
## ENSG00000211951.2 9.24355e-12  
## ENSG00000234184.1 1.16008e-08

head(resSig\_NITvsELI[ order(resSig\_NITvsELI$log2FoldChange, decreasing = TRUE), ])

## log2 fold change (MLE): Group NIT vs ELI   
## Wald test p-value: Group NIT vs ELI   
## DataFrame with 6 rows and 6 columns  
## baseMean log2FoldChange lfcSE stat pvalue  
## <numeric> <numeric> <numeric> <numeric> <numeric>  
## ENSG00000215323.4 4.69040 5.24499 1.919956 2.73183 6.29842e-03  
## ENSG00000110680.8 162.54852 4.15049 1.009889 4.10984 3.95928e-05  
## ENSG00000079689.9 79.87186 3.92513 0.969911 4.04689 5.19020e-05  
## ENSG00000213215.1 1.74278 3.78770 1.123820 3.37038 7.50647e-04  
## ENSG00000163501.6 139.22269 3.64375 0.837913 4.34860 1.37007e-05  
## ENSG00000210184.1 1.89542 3.61009 1.247249 2.89444 3.79838e-03  
## padj  
## <numeric>  
## ENSG00000215323.4 0.041143229  
## ENSG00000110680.8 0.000711661  
## ENSG00000079689.9 0.000894726  
## ENSG00000213215.1 0.007919881  
## ENSG00000163501.6 0.000282809  
## ENSG00000210184.1 0.028257651

write.xlsx(resSig\_NITvsELI, "./Resultados/resSig\_NITvsELI.xlsx")

## SFI vs ELI

res\_SFIvsELI <- results(dds, contrast=c("Group","SFI","ELI"), alpha = 0.05)  
head(res\_SFIvsELI)

## log2 fold change (MLE): Group SFI vs ELI   
## Wald test p-value: Group SFI vs ELI   
## DataFrame with 6 rows and 6 columns  
## baseMean log2FoldChange lfcSE stat pvalue  
## <numeric> <numeric> <numeric> <numeric> <numeric>  
## ENSG00000223972.4 2.841267 0.285092 0.670074 0.425464 0.670498  
## ENSG00000227232.4 711.933057 0.120023 0.235779 0.509047 0.610719  
## ENSG00000243485.2 1.333301 0.710990 0.801901 0.886631 0.375278  
## ENSG00000237613.2 1.653664 0.419062 0.774573 0.541023 0.588492  
## ENSG00000268020.2 0.786348 -1.065733 1.280113 -0.832530 0.405110  
## ENSG00000240361.1 0.706012 -1.421562 1.180453 -1.204251 0.228493  
## padj  
## <numeric>  
## ENSG00000223972.4 0.850645  
## ENSG00000227232.4 0.817353  
## ENSG00000243485.2 NA  
## ENSG00000237613.2 0.803080  
## ENSG00000268020.2 NA  
## ENSG00000240361.1 NA

summary(res\_SFIvsELI)

##   
## out of 35890 with nonzero total read count  
## adjusted p-value < 0.05  
## LFC > 0 (up) : 877, 2.4%  
## LFC < 0 (down) : 2317, 6.5%  
## outliers [1] : 0, 0%  
## low counts [2] : 6963, 19%  
## (mean count < 2)  
## [1] see 'cooksCutoff' argument of ?results  
## [2] see 'independentFiltering' argument of ?results

Se observan 112 genes sobreexpresados y 98 genes downregulados.

Si se considera una fracción de falsos positivos del 10% como aceptable se pueden ordenar mediatne el parámetro de fold change para obtener los genes significates con mayor downregulación y upregulación:

resSig\_SFIvsELI <- subset(res\_SFIvsELI, padj <= 0.05)  
head(resSig\_SFIvsELI[ order(resSig\_SFIvsELI$log2FoldChange), ])

## log2 fold change (MLE): Group SFI vs ELI   
## Wald test p-value: Group SFI vs ELI   
## DataFrame with 6 rows and 6 columns  
## baseMean log2FoldChange lfcSE stat pvalue  
## <numeric> <numeric> <numeric> <numeric> <numeric>  
## ENSG00000160505.11 13.50828 -6.36458 1.379363 -4.61414 3.94721e-06  
## ENSG00000170054.10 39.54938 -6.01153 1.447246 -4.15377 3.27038e-05  
## ENSG00000162897.10 6.85666 -5.59460 1.044443 -5.35654 8.48317e-08  
## ENSG00000221971.3 5.46515 -5.34377 0.890166 -6.00311 1.93568e-09  
## ENSG00000224610.1 2.13161 -5.08906 1.079234 -4.71544 2.41189e-06  
## ENSG00000127074.10 97.23624 -4.90367 0.718360 -6.82620 8.71961e-12  
## padj  
## <numeric>  
## ENSG00000160505.11 2.52960e-04  
## ENSG00000170054.10 1.23358e-03  
## ENSG00000162897.10 1.46087e-05  
## ENSG00000221971.3 1.13380e-06  
## ENSG00000224610.1 1.71025e-04  
## ENSG00000127074.10 2.80297e-08

head(resSig\_SFIvsELI[ order(resSig\_SFIvsELI$log2FoldChange, decreasing = TRUE), ])

## log2 fold change (MLE): Group SFI vs ELI   
## Wald test p-value: Group SFI vs ELI   
## DataFrame with 6 rows and 6 columns  
## baseMean log2FoldChange lfcSE stat pvalue  
## <numeric> <numeric> <numeric> <numeric> <numeric>  
## ENSG00000250033.1 58.74632 4.29926 1.010507 4.25455 2.09468e-05  
## ENSG00000151012.9 2025.87871 4.04723 0.743898 5.44057 5.31092e-08  
## ENSG00000079689.9 79.87186 3.97286 0.969739 4.09683 4.18849e-05  
## ENSG00000174417.2 8.78770 3.73263 0.988974 3.77425 1.60490e-04  
## ENSG00000100604.8 94.12634 3.31623 1.031867 3.21382 1.30982e-03  
## ENSG00000178172.2 1.92728 3.30587 1.091908 3.02761 2.46496e-03  
## padj  
## <numeric>  
## ENSG00000250033.1 0.000879552  
## ENSG00000151012.9 0.000010524  
## ENSG00000079689.9 0.001461801  
## ENSG00000174417.2 0.003975296  
## ENSG00000100604.8 0.018179462  
## ENSG00000178172.2 0.028514146

write.xlsx(resSig\_SFIvsELI, "./Resultados/resSig\_SFIvsELI.xlsx")

# Visualización de patrones de expresión

BiocManager::install("apeglm")

## Bioconductor version 3.11 (BiocManager 1.30.10), R 4.0.0 (2020-04-24)

## Installing package(s) 'apeglm'

## package 'apeglm' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\Judith\AppData\Local\Temp\RtmpKizBus\downloaded\_packages

## Installation path not writeable, unable to update packages: boot, class,  
## foreign, KernSmooth, MASS, nlme, nnet, spatial, survival

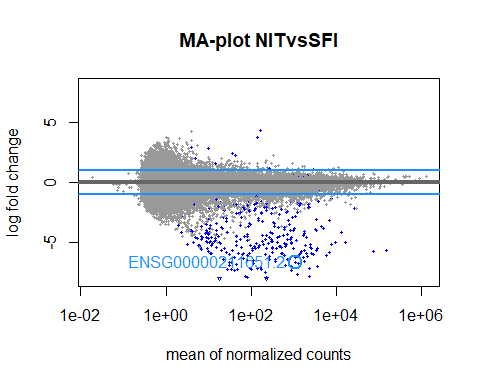
## Old packages: 'limma', 'vctrs'

resultsNames(dds)

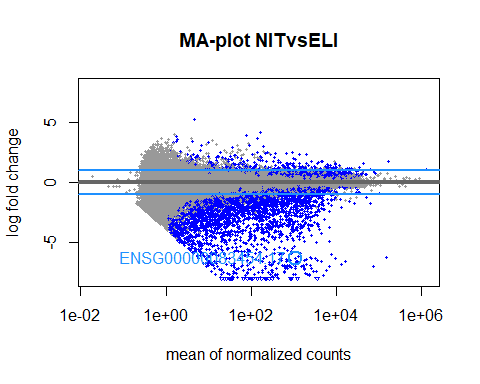
## [1] "Intercept" "Group\_NIT\_vs\_ELI" "Group\_SFI\_vs\_ELI"

## MA-Plot

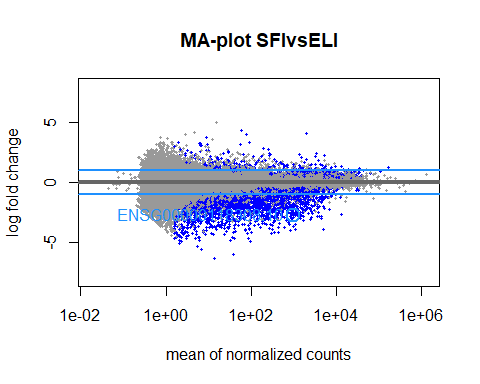
DESeq2::plotMA(res\_NITvsSFI, ylim = c(-8,8))  
topGene1 <- rownames(res\_NITvsSFI)[which.min(res\_NITvsSFI$padj)]  
with(res\_NITvsSFI[topGene1, ], {  
 points(baseMean, log2FoldChange, col = "dodgerblue", cex = 2, lwd = 2)   
 text(baseMean, log2FoldChange, topGene1, pos = 2, col = "dodgerblue")  
})  
abline(h=c(-1,1), col="dodgerblue", lwd=2)  
title("MA-plot NITvsSFI")



DESeq2::plotMA(res\_NITvsELI, ylim = c(-8,8))  
topGene2 <- rownames(res\_NITvsELI)[which.min(res\_NITvsELI$padj)]  
with(res\_NITvsELI[topGene2, ], {  
 points(baseMean, log2FoldChange, col = "dodgerblue", cex = 2, lwd = 2)   
 text(baseMean, log2FoldChange, topGene2, pos = 2, col = "dodgerblue")  
})  
abline(h=c(-1,1), col="dodgerblue", lwd=2)  
title("MA-plot NITvsELI")



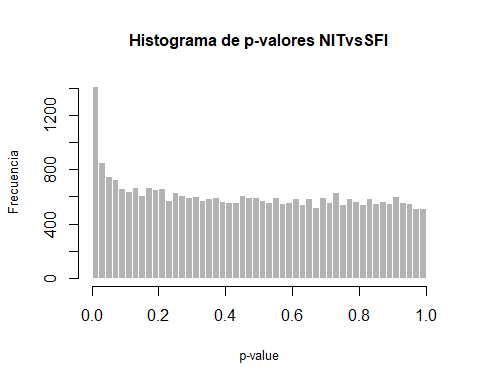
DESeq2::plotMA(res\_SFIvsELI, ylim = c(-8,8))  
topGene3 <- rownames(res\_SFIvsELI)[which.min(res\_SFIvsELI$padj)]  
with(res\_SFIvsELI[topGene3, ], {  
 points(baseMean, log2FoldChange, col = "dodgerblue", cex = 2, lwd = 2)   
 text(baseMean, log2FoldChange, topGene3, pos = 2, col = "dodgerblue")  
})  
abline(h=c(-1,1), col="dodgerblue", lwd=2)  
title("MA-plot SFIvsELI")



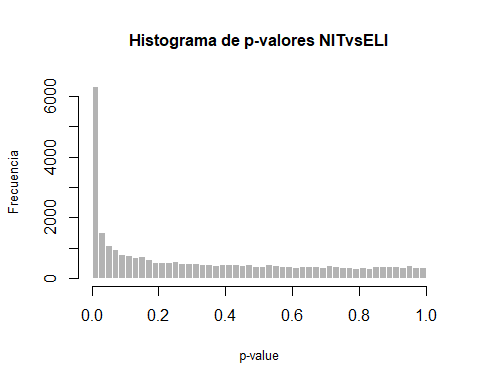
**Mirar si fer amb lfcShrink:**

## Histograma de los p-valores

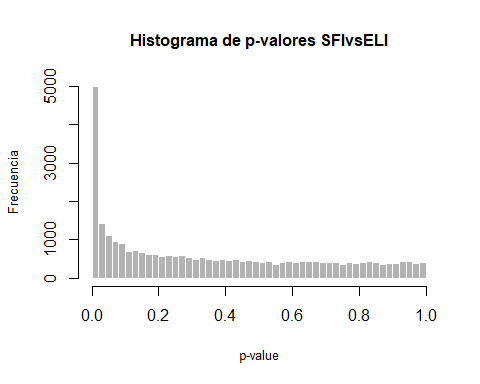
hist(res\_NITvsSFI$pvalue[res\_NITvsSFI$baseMean > 1], breaks = 0:50/50, col = "grey70", border = "white", main = "Histograma de p-valores NITvsSFI", cex.main = 1, ylab = "Frecuencia", xlab = "p-value", cex.lab = 0.8)

 Mirar teoria DESEQ2 from Michael I. Love.

hist(res\_NITvsELI$pvalue[res\_NITvsELI$baseMean > 1], breaks = 0:50/50, col = "grey70", border = "white", main = "Histograma de p-valores NITvsELI", cex.main = 1, ylab = "Frecuencia", xlab = "p-value", cex.lab = 0.8)

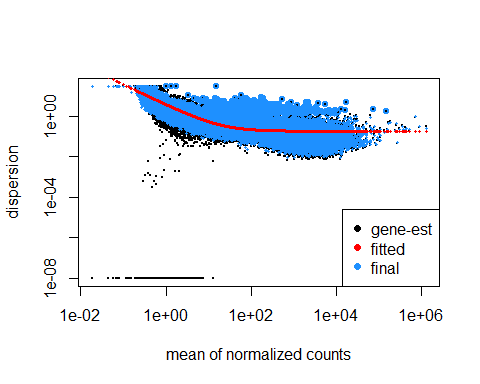


hist(res\_SFIvsELI$pvalue[res\_SFIvsELI$baseMean > 1], breaks = 0:50/50, col = "grey70", border = "white", main = "Histograma de p-valores SFIvsELI", cex.main = 1, ylab = "Frecuencia", xlab = "p-value", cex.lab = 0.8)



## Dispersion plot and fitting alternatives

plotDispEsts(dds)



## Gene clustering

library("genefilter")

##   
## Attaching package: 'genefilter'

## The following objects are masked from 'package:matrixStats':  
##   
## rowSds, rowVars

topVarGenes <- head(order(rowVars(assay(vsd)), decreasing = TRUE), 20)  
topVarGenes

## [1] 35151 24617 35801 35849 35828 24614 24618 35826 35846 24609 4539 33862  
## [13] 33899 19750 24748 4517 4520 4524 24608 24603

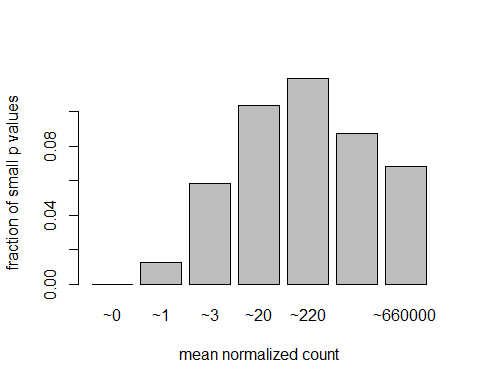
mat <- assay(vsd)[topVarGenes, ]  
mat <- mat - rowMeans(mat)  
anno <- as.data.frame(colData(vsd)["Group"])  
pdf("pheatmap2.pdf")  
pheatmap2 <- pheatmap(mat, annotation\_col = anno)  
dev.off()

## pdf   
## 3

## Independent filtering

Una de las debilidades de los datos RNAseq es que los genes poco expresados no es posible ver expresion diferencial, debido al ruido de fondo.

qs <- c(0, quantile(res\_NITvsSFI$baseMean[res\_NITvsSFI$baseMean > 0], 0:6/6))  
bins <- cut(res\_NITvsSFI$baseMean, qs)  
levels(bins) <- paste0("~", round(signif((qs[-1] + qs[-length(qs)])/2, 2)))  
fractionSig <- tapply(res\_NITvsSFI$pvalue, bins, function(p)  
 mean(p < .05, na.rm = TRUE))  
barplot(fractionSig, xlab = "mean normalized count",  
 ylab = "fraction of small p values")



Este gráfico muestra que los genes con contajes muy bajos no tienen poder y estan excluidos del test. Estos genes tienen cierta influencia en el ajuste de testado múltiple, y su ejecución mejora cuando estos genes son excluidos.

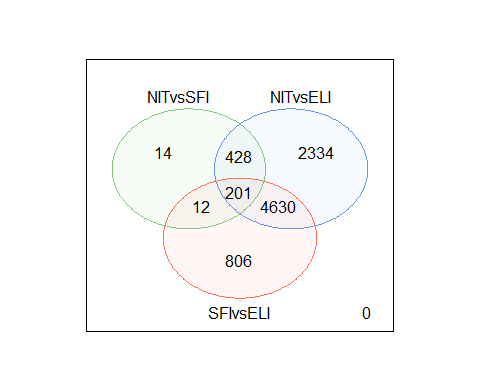
## Comparación entre distintas comparaciones

res\_NITvsSFI.genes <- row.names(resSig\_NITvsSFI)  
res\_NITvsELI.genes <- row.names(resSig\_NITvsELI)  
res\_SFIvsELI.genes <- row.names(resSig\_SFIvsELI)

comb <- c(res\_NITvsSFI.genes,res\_NITvsELI.genes,res\_SFIvsELI.genes)

res\_NITvsSFI.genes.2 <- comb %in% res\_NITvsSFI.genes  
res\_NITvsELI.genes.2 <- comb %in% res\_NITvsELI.genes  
res\_SFIvsELI.genes.2 <- comb %in% res\_SFIvsELI.genes

venn\_counts <- cbind(res\_NITvsSFI.genes.2, res\_NITvsELI.genes.2, res\_SFIvsELI.genes.2)  
venn\_counts\_results <- vennCounts(venn\_counts)  
vennDiagram(venn\_counts\_results, cex = 1, names=c("NITvsSFI","NITvsELI","SFIvsELI"), circle.col=c("palegreen3","cornflowerblue","coral1"))



# Anotación de los resultados

BiocManager::install("AnnotationDbi")

## Bioconductor version 3.11 (BiocManager 1.30.10), R 4.0.0 (2020-04-24)

## Installing package(s) 'AnnotationDbi'

## package 'AnnotationDbi' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\Judith\AppData\Local\Temp\RtmpKizBus\downloaded\_packages

## Installation path not writeable, unable to update packages: boot, class,  
## foreign, KernSmooth, MASS, nlme, nnet, spatial, survival

## Old packages: 'limma', 'vctrs'

BiocManager::install("org.Hs.eg.db")

## Bioconductor version 3.11 (BiocManager 1.30.10), R 4.0.0 (2020-04-24)

## Installing package(s) 'org.Hs.eg.db'

## installing the source package 'org.Hs.eg.db'

## Warning in install.packages(...): installation of package 'org.Hs.eg.db' had  
## non-zero exit status

## Installation path not writeable, unable to update packages: boot, class,  
## foreign, KernSmooth, MASS, nlme, nnet, spatial, survival

## Old packages: 'limma', 'vctrs'

library("AnnotationDbi")

##   
## Attaching package: 'AnnotationDbi'

## The following object is masked from 'package:dplyr':  
##   
## select

library("org.Hs.eg.db")

##

columns(org.Hs.eg.db)

## [1] "ACCNUM" "ALIAS" "ENSEMBL" "ENSEMBLPROT" "ENSEMBLTRANS"  
## [6] "ENTREZID" "ENZYME" "EVIDENCE" "EVIDENCEALL" "GENENAME"   
## [11] "GO" "GOALL" "IPI" "MAP" "OMIM"   
## [16] "ONTOLOGY" "ONTOLOGYALL" "PATH" "PFAM" "PMID"   
## [21] "PROSITE" "REFSEQ" "SYMBOL" "UCSCKG" "UNIGENE"   
## [26] "UNIPROT"

To remove the string after the period:

row.names(res\_NITvsSFI) <- gsub(x=row.names(res\_NITvsSFI), pattern = "\\..\*", replacement = "")  
row.names(res\_NITvsELI) <- gsub(x=row.names(res\_NITvsELI), pattern = "\\..\*", replacement = "")  
row.names(res\_SFIvsELI) <- gsub(x=row.names(res\_SFIvsELI), pattern = "\\..\*", replacement = "")

res\_NITvsSFI$symbol <- mapIds(org.Hs.eg.db, keys=row.names(res\_NITvsSFI), column = "SYMBOL", keytype="ENSEMBL", multiVals = "first")

## 'select()' returned 1:many mapping between keys and columns

res\_NITvsSFI$entrez <- mapIds(org.Hs.eg.db, keys=row.names(res\_NITvsSFI), column = "ENTREZID", keytype="ENSEMBL", multiVals = "first")

## 'select()' returned 1:many mapping between keys and columns

res\_NITvsSFI$genename <- mapIds(org.Hs.eg.db, keys=row.names(res\_NITvsSFI), column = "GENENAME", keytype="ENSEMBL", multiVals = "first")

## 'select()' returned 1:many mapping between keys and columns

res\_NITvsSFI\_Ordered <- res\_NITvsSFI[order(res\_NITvsSFI$pvalue),]  
res\_NITvsSFI\_Annot <- res\_NITvsSFI\_Ordered[which(res\_NITvsSFI\_Ordered$symbol != "NA"), ]  
head(res\_NITvsSFI\_Annot)

## log2 fold change (MLE): Group NIT vs SFI   
## Wald test p-value: Group NIT vs SFI   
## DataFrame with 6 rows and 9 columns  
## baseMean log2FoldChange lfcSE stat pvalue  
## <numeric> <numeric> <numeric> <numeric> <numeric>  
## ENSG00000132465 7719.061 -5.14045 0.801605 -6.41270 1.42962e-10  
## ENSG00000105369 1872.744 -5.28248 0.845055 -6.25104 4.07719e-10  
## ENSG00000170476 1195.810 -4.86043 0.784789 -6.19329 5.89206e-10  
## ENSG00000110777 2016.965 -4.59036 0.791125 -5.80232 6.54051e-09  
## ENSG00000117322 871.678 -5.07437 0.952255 -5.32880 9.88654e-08  
## ENSG00000143297 1677.871 -4.39256 0.830238 -5.29073 1.21832e-07  
## padj symbol entrez  
## <numeric> <character> <character>  
## ENSG00000132465 9.84557e-08 JCHAIN 3512  
## ENSG00000105369 2.55830e-07 CD79A 973  
## ENSG00000170476 3.32737e-07 MZB1 51237  
## ENSG00000110777 3.07796e-06 POU2AF1 5450  
## ENSG00000117322 3.36333e-05 CR2 1380  
## ENSG00000143297 3.95409e-05 FCRL5 83416  
## genename  
## <character>  
## ENSG00000132465 joining chain of multimeric IgA and IgM  
## ENSG00000105369 CD79a molecule  
## ENSG00000170476 marginal zone B and B1 cell specific protein  
## ENSG00000110777 POU class 2 homeobox associating factor 1  
## ENSG00000117322 complement C3d receptor 2  
## ENSG00000143297 Fc receptor like 5

res\_NITvsELI$symbol <- mapIds(org.Hs.eg.db, keys=row.names(res\_NITvsELI), column = "SYMBOL", keytype="ENSEMBL", multiVals = "first")

## 'select()' returned 1:many mapping between keys and columns

res\_NITvsELI$entrez <- mapIds(org.Hs.eg.db, keys=row.names(res\_NITvsELI), column = "ENTREZID", keytype="ENSEMBL", multiVals = "first")

## 'select()' returned 1:many mapping between keys and columns

res\_NITvsELI$genename <- mapIds(org.Hs.eg.db, keys=row.names(res\_NITvsELI), column = "GENENAME", keytype="ENSEMBL", multiVals = "first")

## 'select()' returned 1:many mapping between keys and columns

res\_NITvsELI\_Ordered <- res\_NITvsELI[order(res\_NITvsELI$pvalue),]  
res\_NITvsELI\_Annot <- res\_NITvsELI\_Ordered[which(res\_NITvsELI\_Ordered$symbol != "NA"), ]  
head(res\_NITvsELI\_Annot)

## log2 fold change (MLE): Group NIT vs ELI   
## Wald test p-value: Group NIT vs ELI   
## DataFrame with 6 rows and 9 columns  
## baseMean log2FoldChange lfcSE stat pvalue  
## <numeric> <numeric> <numeric> <numeric> <numeric>  
## ENSG00000083454 1101.334 -6.31926 0.623398 -10.13680 3.79338e-24  
## ENSG00000136573 1092.603 -6.79868 0.694097 -9.79501 1.18290e-22  
## ENSG00000167483 809.646 -6.94637 0.709558 -9.78971 1.24649e-22  
## ENSG00000156738 4902.360 -7.76238 0.808529 -9.60062 7.94686e-22  
## ENSG00000035720 148.888 -6.48581 0.679336 -9.54727 1.33156e-21  
## ENSG00000173200 1732.070 -4.59398 0.483754 -9.49652 2.17016e-21  
## padj symbol entrez  
## <numeric> <character> <character>  
## ENSG00000083454 1.15027e-19 P2RX5 5026  
## ENSG00000136573 1.25991e-18 BLK 640  
## ENSG00000167483 1.25991e-18 NIBAN3 199786  
## ENSG00000156738 6.02432e-18 MS4A1 931  
## ENSG00000035720 7.86046e-18 STAP1 26228  
## ENSG00000173200 9.40083e-18 PARP15 165631  
## genename  
## <character>  
## ENSG00000083454 purinergic receptor P2X 5  
## ENSG00000136573 BLK proto-oncogene, Src family tyrosine kinase  
## ENSG00000167483 niban apoptosis regulator 3  
## ENSG00000156738 membrane spanning 4-domains A1  
## ENSG00000035720 signal transducing adaptor family member 1  
## ENSG00000173200 poly(ADP-ribose) polymerase family member 15

res\_SFIvsELI$symbol <- mapIds(org.Hs.eg.db, keys=row.names(res\_SFIvsELI), column = "SYMBOL", keytype="ENSEMBL", multiVals = "first")

## 'select()' returned 1:many mapping between keys and columns

res\_SFIvsELI$entrez <- mapIds(org.Hs.eg.db, keys=row.names(res\_SFIvsELI), column = "ENTREZID", keytype="ENSEMBL", multiVals = "first")

## 'select()' returned 1:many mapping between keys and columns

res\_SFIvsELI$genename <- mapIds(org.Hs.eg.db, keys=row.names(res\_SFIvsELI), column = "GENENAME", keytype="ENSEMBL", multiVals = "first")

## 'select()' returned 1:many mapping between keys and columns

res\_SFIvsELI\_Ordered <- res\_SFIvsELI[order(res\_SFIvsELI$pvalue),]  
res\_SFIvsELI\_Annot <- res\_SFIvsELI\_Ordered[which(res\_SFIvsELI\_Ordered$symbol != "NA"), ]  
head(res\_SFIvsELI\_Annot)

## log2 fold change (MLE): Group SFI vs ELI   
## Wald test p-value: Group SFI vs ELI   
## DataFrame with 6 rows and 9 columns  
## baseMean log2FoldChange lfcSE stat pvalue  
## <numeric> <numeric> <numeric> <numeric> <numeric>  
## ENSG00000118308 984.664 -2.76216 0.360386 -7.66444 1.79612e-14  
## ENSG00000069493 1084.588 -2.25105 0.308833 -7.28887 3.12568e-13  
## ENSG00000074966 420.444 -3.02665 0.424304 -7.13323 9.80417e-13  
## ENSG00000196172 220.306 -1.16607 0.167624 -6.95644 3.48973e-12  
## ENSG00000111913 1434.234 -2.47621 0.356321 -6.94938 3.66890e-12  
## ENSG00000068831 3683.217 -2.87925 0.415866 -6.92352 4.40562e-12  
## padj symbol entrez  
## <numeric> <character> <character>  
## ENSG00000118308 5.19634e-10 LRMP 4033  
## ENSG00000069493 4.52145e-09 CLEC2D 29121  
## ENSG00000074966 9.45481e-09 TXK 7294  
## ENSG00000196172 2.02652e-08 ZNF681 148213  
## ENSG00000111913 2.02652e-08 RIPOR2 9750  
## ENSG00000068831 2.02652e-08 RASGRP2 10235  
## genename  
## <character>  
## ENSG00000118308 lymphoid restricted membrane protein  
## ENSG00000069493 C-type lectin domain family 2 member D  
## ENSG00000074966 TXK tyrosine kinase  
## ENSG00000196172 zinc finger protein 681  
## ENSG00000111913 RHO family interacting cell polarization regulator 2  
## ENSG00000068831 RAS guanyl releasing protein 2

write.xlsx(res\_SFIvsELI\_Annot, "./Resultados/res\_SFIvsELI\_Annot.xlsx")

## Volcano plots:

genesymbols1 <- res\_NITvsSFI\_Annot$symbol  
pdf("volcano\_NITvsSFI.pdf")  
volcano\_NITvsSFI <- plot(res\_NITvsSFI\_Annot$log2FoldChange, -log10(res\_NITvsSFI\_Annot$padj), panel.first = grid(), main = "Volcano plot NITvsSFI", xlab="Effect size: log2 Fold-change", ylab="-log10(adj p-value)", pch=20, cex=0.6)  
abline(v=0)  
abline(v=c(-1,1), col="dodgerblue")  
abline(h=-log10(0.05), col="dodgerblue")  
gn.selected <- abs(res\_NITvsSFI\_Annot$log2FoldChange) > 2.5 & res\_NITvsSFI\_Annot$padj < 0.05   
text(res\_NITvsSFI\_Annot$log2FoldChange[gn.selected],  
 -log10(res\_NITvsSFI\_Annot$padj)[gn.selected],  
 lab=genesymbols1[gn.selected], cex=0.4)  
dev.off()

## png   
## 2

genesymbols2 <- res\_NITvsELI\_Annot$symbol  
pdf("volcano\_NITvsELI.pdf")  
volano\_NITvsELI <- plot(res\_NITvsELI\_Annot$log2FoldChange, -log10(res\_NITvsELI\_Annot$padj), panel.first = grid(), main = "Volcano plot NITvsELI", xlab="Effect size: log2 Fold-change", ylab="-log10(adj p-value)", pch=20, cex=0.6)  
abline(v=0)  
abline(v=c(-1,1), col="dodgerblue")  
abline(h=-log10(0.05), col="dodgerblue")  
gn.selected <- abs(res\_NITvsELI\_Annot$log2FoldChange) > 4 & res\_NITvsELI\_Annot$padj < 0.05   
text(res\_NITvsELI\_Annot$log2FoldChange[gn.selected],  
 -log10(res\_NITvsELI\_Annot$padj)[gn.selected],  
 lab=genesymbols2[gn.selected], cex=0.4)  
dev.off()

## png   
## 2

genesymbols3 <- res\_SFIvsELI\_Annot$symbol  
pdf("volcano\_SFIvsELI.pdf")  
volcano\_SFIvsELI <- plot(res\_SFIvsELI\_Annot$log2FoldChange, -log10(res\_SFIvsELI\_Annot$padj), panel.first = grid(), main = "Volcano plot SFIvsELI", xlab="Effect size: log2 Fold-change", ylab="-log10(adj p-value)", pch=20, cex=0.6)  
abline(v=0)  
abline(v=c(-1,1), col="dodgerblue")  
abline(h=-log10(0.05), col="dodgerblue")  
gn.selected <- abs(res\_SFIvsELI\_Annot$log2FoldChange) > 4 & res\_SFIvsELI\_Annot$padj < 0.05   
text(res\_SFIvsELI\_Annot$log2FoldChange[gn.selected],  
 -log10(res\_SFIvsELI\_Annot$padj)[gn.selected],  
 lab=genesymbols3[gn.selected], cex=0.4)  
dev.off()

## png   
## 2

# Análisis de significación biológica

BiocManager::install("clusterProfiler")

## Bioconductor version 3.11 (BiocManager 1.30.10), R 4.0.0 (2020-04-24)

## Installing package(s) 'clusterProfiler'

## package 'clusterProfiler' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\Judith\AppData\Local\Temp\RtmpKizBus\downloaded\_packages

## Installation path not writeable, unable to update packages: boot, class,  
## foreign, KernSmooth, MASS, nlme, nnet, spatial, survival

## Old packages: 'limma', 'vctrs'

BiocManager::install("xlsx")

## Bioconductor version 3.11 (BiocManager 1.30.10), R 4.0.0 (2020-04-24)

## Installing package(s) 'xlsx'

## Warning: package 'xlsx' is in use and will not be installed

## Installation path not writeable, unable to update packages: boot, class,  
## foreign, KernSmooth, MASS, nlme, nnet, spatial, survival

## Old packages: 'limma', 'vctrs'

library(clusterProfiler)

##

## clusterProfiler v3.16.0 For help: https://guangchuangyu.github.io/software/clusterProfiler  
##   
## If you use clusterProfiler in published research, please cite:  
## Guangchuang Yu, Li-Gen Wang, Yanyan Han, Qing-Yu He. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS: A Journal of Integrative Biology. 2012, 16(5):284-287.

##   
## Attaching package: 'clusterProfiler'

## The following object is masked from 'package:AnnotationDbi':  
##   
## select

## The following object is masked from 'package:DelayedArray':  
##   
## simplify

## The following object is masked from 'package:IRanges':  
##   
## slice

## The following object is masked from 'package:S4Vectors':  
##   
## rename

## The following object is masked from 'package:stats':  
##   
## filter

GO\_NITvsSFI <- as.data.frame(res\_NITvsSFI\_Annot)  
head(GO\_NITvsSFI)

## baseMean log2FoldChange lfcSE stat pvalue  
## ENSG00000132465 7719.0610 -5.140455 0.8016050 -6.412703 1.429622e-10  
## ENSG00000105369 1872.7436 -5.282475 0.8450550 -6.251044 4.077190e-10  
## ENSG00000170476 1195.8098 -4.860429 0.7847893 -6.193291 5.892063e-10  
## ENSG00000110777 2016.9654 -4.590357 0.7911249 -5.802316 6.540506e-09  
## ENSG00000117322 871.6781 -5.074373 0.9522549 -5.328797 9.886536e-08  
## ENSG00000143297 1677.8711 -4.392559 0.8302376 -5.290725 1.218322e-07  
## padj symbol entrez  
## ENSG00000132465 9.845565e-08 JCHAIN 3512  
## ENSG00000105369 2.558301e-07 CD79A 973  
## ENSG00000170476 3.327366e-07 MZB1 51237  
## ENSG00000110777 3.077962e-06 POU2AF1 5450  
## ENSG00000117322 3.363328e-05 CR2 1380  
## ENSG00000143297 3.954086e-05 FCRL5 83416  
## genename  
## ENSG00000132465 joining chain of multimeric IgA and IgM  
## ENSG00000105369 CD79a molecule  
## ENSG00000170476 marginal zone B and B1 cell specific protein  
## ENSG00000110777 POU class 2 homeobox associating factor 1  
## ENSG00000117322 complement C3d receptor 2  
## ENSG00000143297 Fc receptor like 5

GO\_NITvsELI <- as.data.frame(res\_NITvsELI\_Annot)  
head(GO\_NITvsELI)

## baseMean log2FoldChange lfcSE stat pvalue  
## ENSG00000083454 1101.3342 -6.319262 0.6233983 -10.136796 3.793380e-24  
## ENSG00000136573 1092.6033 -6.798680 0.6940966 -9.795007 1.182897e-22  
## ENSG00000167483 809.6461 -6.946372 0.7095583 -9.789713 1.246489e-22  
## ENSG00000156738 4902.3599 -7.762379 0.8085294 -9.600615 7.946862e-22  
## ENSG00000035720 148.8879 -6.485810 0.6793364 -9.547273 1.331556e-21  
## ENSG00000173200 1732.0704 -4.593980 0.4837539 -9.496523 2.170161e-21  
## padj symbol entrez  
## ENSG00000083454 1.150267e-19 P2RX5 5026  
## ENSG00000136573 1.259910e-18 BLK 640  
## ENSG00000167483 1.259910e-18 NIBAN3 199786  
## ENSG00000156738 6.024317e-18 MS4A1 931  
## ENSG00000035720 7.860456e-18 STAP1 26228  
## ENSG00000173200 9.400827e-18 PARP15 165631  
## genename  
## ENSG00000083454 purinergic receptor P2X 5  
## ENSG00000136573 BLK proto-oncogene, Src family tyrosine kinase  
## ENSG00000167483 niban apoptosis regulator 3  
## ENSG00000156738 membrane spanning 4-domains A1  
## ENSG00000035720 signal transducing adaptor family member 1  
## ENSG00000173200 poly(ADP-ribose) polymerase family member 15

GO\_SFIvsELI <- as.data.frame(res\_SFIvsELI\_Annot)  
head(GO\_SFIvsELI)

## baseMean log2FoldChange lfcSE stat pvalue  
## ENSG00000118308 984.6640 -2.762156 0.3603859 -7.664441 1.796116e-14  
## ENSG00000069493 1084.5876 -2.251046 0.3088334 -7.288869 3.125681e-13  
## ENSG00000074966 420.4444 -3.026654 0.4243036 -7.133228 9.804170e-13  
## ENSG00000196172 220.3064 -1.166066 0.1676239 -6.956442 3.489729e-12  
## ENSG00000111913 1434.2339 -2.476210 0.3563209 -6.949382 3.668903e-12  
## ENSG00000068831 3683.2166 -2.879255 0.4158658 -6.923518 4.405620e-12  
## padj symbol entrez  
## ENSG00000118308 5.196342e-10 LRMP 4033  
## ENSG00000069493 4.521455e-09 CLEC2D 29121  
## ENSG00000074966 9.454815e-09 TXK 7294  
## ENSG00000196172 2.026521e-08 ZNF681 148213  
## ENSG00000111913 2.026521e-08 RIPOR2 9750  
## ENSG00000068831 2.026521e-08 RASGRP2 10235  
## genename  
## ENSG00000118308 lymphoid restricted membrane protein  
## ENSG00000069493 C-type lectin domain family 2 member D  
## ENSG00000074966 TXK tyrosine kinase  
## ENSG00000196172 zinc finger protein 681  
## ENSG00000111913 RHO family interacting cell polarization regulator 2  
## ENSG00000068831 RAS guanyl releasing protein 2

library("dplyr")  
library("xlsx")  
universe\_prova <- list(NITvsSFI = GO\_NITvsSFI$entrez, NITvsELI = GO\_NITvsELI$entrez, SFIvsELI = GO\_SFIvsELI$entrez)  
func <- function(x) {  
 x %>% filter(padj < 0.05, !is.na(entrez)) %>% pull(entrez)  
}  
sigGenes\_prova <- list(NITvsSFI = GO\_NITvsSFI, NITvsELI = GO\_NITvsELI, SFIvsELI = GO\_SFIvsELI) %>% lapply(func)  
comparisonsNames <- names(sigGenes\_prova)  
  
for (i in 1:length(sigGenes\_prova)){  
 genesIn <- sigGenes\_prova[[i]]  
 comparison <- comparisonsNames[i]  
 enrich.result <- enrichGO(gene = genesIn, OrgDb = org.Hs.eg.db, ont = "ALL", pAdjustMethod = "BH", pvalueCutoff = 0.05, universe = universe\_prova, readable = TRUE)  
  
cat("##################################")  
 cat("\nComparison: ", comparison,"\n")  
 print(head(enrich.result))  
  
if (length(rownames(enrich.result@result)) != 0) {  
 write.csv(as.data.frame(enrich.result),   
 file =paste0("./Resultados/","Enrich.Results.",comparison,".csv"),   
 row.names = FALSE)  
  
write.xlsx(as.data.frame(enrich.result),   
 file =paste0("./Resultados/","Enrich.Results.",comparison,".xlsx"),   
 row.names = FALSE)   
  
 pdf(file=paste0("./Resultados/","Enrich.Dotplot.",comparison,".pdf"))  
 print(dotplot(enrich.result, showCategory = 15, font.size = 6,   
 title = paste0("EnrichGO Pathway Analysis for ", comparison,". Dotplot")))  
 dev.off()  
  
 pdf(file = paste0("./Resultados/","EnrichGOemapplot.",comparison,".pdf"))  
 print(emapplot(enrich.result, categorySize = "geneNum", schowCategory = 15,   
 vertex.label.cex = 0.75))  
 dev.off()  
  
 }  
}

## ##################################  
## Comparison: NITvsSFI   
## ONTOLOGY ID Description  
## GO:0030098 BP GO:0030098 lymphocyte differentiation  
## GO:0042113 BP GO:0042113 B cell activation  
## GO:0042100 BP GO:0042100 B cell proliferation  
## GO:0050853 BP GO:0050853 B cell receptor signaling pathway  
## GO:0070661 BP GO:0070661 leukocyte proliferation  
## GO:0050851 BP GO:0050851 antigen receptor-mediated signaling pathway  
## GeneRatio BgRatio pvalue p.adjust qvalue  
## GO:0030098 16/83 353/18670 3.117187e-12 4.809819e-09 4.012968e-09  
## GO:0042113 15/83 310/18670 6.316444e-12 4.873137e-09 4.065796e-09  
## GO:0042100 10/83 95/18670 1.283210e-11 6.599978e-09 5.506548e-09  
## GO:0050853 10/83 129/18670 2.765868e-10 1.066934e-07 8.901729e-08  
## GO:0070661 13/83 298/18670 6.545921e-10 2.020071e-07 1.685402e-07  
## GO:0050851 13/83 316/18670 1.337265e-09 3.439000e-07 2.869255e-07  
## geneID  
## GO:0030098 CD79A/CR2/MS4A1/LEPR/IRF4/FCRL3/CD19/LY9/IHH/TNFRSF18/ITK/RUNX3/CD28/PTPRC/EOMES/IKZF3  
## GO:0042113 CD79A/MZB1/CR2/TNFRSF13B/MS4A1/IGLL5/FCRL3/CD19/FCRL1/SLA2/CXCR5/TNFRSF13C/CD28/PTPRC/IKZF3  
## GO:0042100 CD79A/MZB1/CR2/TNFRSF13B/MS4A1/FCRL3/CD19/TNFRSF13C/PTPRC/IKZF3  
## GO:0050853 CD79A/MS4A1/IGLL5/FCRL3/CD19/STAP1/BLK/PAX5/ITK/PTPRC  
## GO:0070661 CD79A/MZB1/CR2/TNFRSF13B/MS4A1/FCRL3/CD19/IL5RA/IHH/TNFRSF13C/CD28/PTPRC/IKZF3  
## GO:0050851 CD79A/MS4A1/IGLL5/FCRL3/CD19/STAP1/BLK/TRAT1/SLA2/PAX5/ITK/CD28/PTPRC  
## Count  
## GO:0030098 16  
## GO:0042113 15  
## GO:0042100 10  
## GO:0050853 10  
## GO:0070661 13  
## GO:0050851 13

## ##################################  
## Comparison: NITvsELI   
## ONTOLOGY ID Description GeneRatio  
## GO:0042110 BP GO:0042110 T cell activation 196/3069  
## GO:0030098 BP GO:0030098 lymphocyte differentiation 153/3069  
## GO:0050863 BP GO:0050863 regulation of T cell activation 141/3069  
## GO:0051249 BP GO:0051249 regulation of lymphocyte activation 180/3069  
## GO:0007159 BP GO:0007159 leukocyte cell-cell adhesion 140/3069  
## GO:0022407 BP GO:0022407 regulation of cell-cell adhesion 156/3069  
## BgRatio pvalue p.adjust qvalue  
## GO:0042110 464/18670 1.266685e-40 7.938313e-37 6.385424e-37  
## GO:0030098 353/18670 2.022998e-33 6.339064e-30 5.099020e-30  
## GO:0050863 314/18670 6.935144e-33 1.448751e-29 1.165347e-29  
## GO:0051249 485/18670 7.228436e-29 1.132515e-25 9.109732e-26  
## GO:0007159 337/18670 2.306607e-28 2.891101e-25 2.325545e-25  
## GO:0022407 402/18670 1.432641e-27 1.496394e-24 1.203670e-24  
## geneID  
## GO:0042110 ITK/CTLA4/BTLA/SLAMF6/CAMK4/IRF4/TIGIT/LAX1/SIRPG/KLRK1/PTPRC/CLECL1/CARD11/PTPN22/LY9/RIPOR2/CD6/CD28/BCL11B/RUNX3/RHOH/SLA2/RASAL3/LCK/GRAP2/PDCD1/RASGRP1/CCDC88B/CD3G/NLRC3/ZAP70/ITGAL/CD3E/TNFRSF13C/TESPA1/EOMES/SIT1/DOCK2/WAS/DOCK8/MYB/IL7R/CD2/GPR18/TCF7/CD3D/CORO1A/TBX21/SASH3/CD5/RAC2/HLA-DPB1/SPN/IL12RB1/LAG3/CD40LG/LILRB1/TNFSF8/JAML/ICOS/VAV1/FANCD2/TREML2/THEMIS/NOD2/FOXP3/HLA-DOA/BTN2A2/CCR6/BTN3A1/CD27/TNFRSF18/CD8A/JAK3/CD7/FCGR2B/APBB1IP/LCP1/CD1C/WNT1/PIK3CD/ITPKB/NCKAP1L/PTPN6/FUT7/CASP8/IL4R/CD300A/CR1/P2RX7/F2RL1/CCR7/HLA-DPA1/PIK3CG/SIRPB1/TNFSF4/RUNX2/CD83/CD74/NFATC2/PAG1/CD8B/BATF/ADAM8/TMEM131L/MALT1/CD70/GPR183/IL23A/TNFSF11/CD86/HAVCR2/ABL1/TMIGD2/TNFRSF1B/IL18/CSK/ZMIZ1/CCR2/ERBB2/VNN1/NLRP3/IHH/HHLA2/AIRE/PLA2G2D/IL18R1/DPP4/ELF4/CD47/PELI1/CD274/TNFSF13B/B2M/CASP3/FOXN1/IFNG/LAT/FANCA/TCIRG1/PRDM1/IL23R/LGALS9B/IL7/HLA-G/MICB/FAM49B/ZBTB7B/ADORA2A/PYCARD/CCR9/EFNB1/CLEC7A/CD4/HLA-DMB/TNFAIP8L2/LGALS9/IL12A/IL2/FYN/SMAD7/CD80/LEF1/SELENOK/NCSTN/KDELR1/IL2RA/LFNG/GLI3/EGR3/IL10/AP3D1/DDOST/AP3B1/LEPR/DUSP22/ZBTB1/ZC3H8/PATZ1/CGAS/CD160/FZD5/HMGB1/EFNB2/CLPTM1/LILRB4/RAC1/SH3RF1/HLA-E/PTPN2/XCL1/VCAM1/CYLD/STOML2/FKBP1A/PSMB10  
## GO:0030098 MS4A1/FCRL3/ITK/CTLA4/CD79A/CD19/SLAMF6/CAMK4/IRF4/CR2/PTPRC/POU2F2/CD79B/CARD11/PTPN22/IKZF3/IKZF1/LY9/INPP5D/CD28/BCL11B/RUNX3/RHOH/AICDA/HDAC9/LCK/RASGRP1/CD3G/ZAP70/ITGA4/CD3E/TESPA1/EOMES/DOCK2/MYB/IL7R/CD2/GPR18/TCF7/CD3D/TBX21/SASH3/DOCK10/SPN/IL12RB1/LAG3/PTK2B/BTK/CD40LG/PLCG2/TNFSF8/VAV1/FANCD2/THEMIS/FOXP3/HLA-DOA/BTN2A2/ATM/CCR6/TOX/CD27/TNFRSF18/CD8A/JAK3/FCGR2B/WNT1/PIK3CD/ITPKB/NCKAP1L/PTPN6/FUT7/IL4R/CR1/CCR7/TNFSF4/RUNX2/CD83/CD74/NFATC2/BATF/ADAM8/TMEM131L/NFAM1/MALT1/C17orf99/GPR183/IL23A/FLT3/CD86/ABL1/IL18/ZMIZ1/CCR2/ERBB2/VNN1/HMGB3/NLRP3/IHH/AIRE/PLA2G2D/IL18R1/B2M/FOXN1/IFNG/FANCA/TCIRG1/SPI1/PRDM1/IL23R/IL7/SLAMF8/HLA-G/ZBTB7B/CCR9/TLR9/PGLYRP2/LYL1/CD4/LGALS9/IL12A/IL2/SMAD7/CD80/LEF1/ATP11C/ZBTB7A/KDELR1/IL2RA/BLNK/LFNG/GLI3/EGR3/IL10/AP3D1/AP3B1/LEPR/ZBTB1/ZC3H8/PATZ1/FZD5/PLCL2/HMGB1/CLPTM1/LILRB4/PPP2R3C/SH3RF1/CDH17/FLT3LG/PTPN2/VCAM1/RBPJ/CYLD/ITGB1  
## GO:0050863 CTLA4/BTLA/CAMK4/IRF4/TIGIT/LAX1/SIRPG/KLRK1/PTPRC/CLECL1/CARD11/PTPN22/RIPOR2/CD6/CD28/RUNX3/RASAL3/LCK/GRAP2/PDCD1/RASGRP1/CCDC88B/ZAP70/CD3E/TNFRSF13C/TESPA1/SIT1/DOCK8/MYB/IL7R/CD2/CORO1A/TBX21/SASH3/CD5/RAC2/HLA-DPB1/SPN/IL12RB1/LAG3/CD40LG/LILRB1/TNFSF8/ICOS/VAV1/FANCD2/NOD2/FOXP3/HLA-DOA/BTN2A2/CD27/TNFRSF18/JAK3/FCGR2B/ITPKB/NCKAP1L/PTPN6/IL4R/CD300A/CR1/CCR7/HLA-DPA1/SIRPB1/TNFSF4/CD83/CD74/NFATC2/PAG1/ADAM8/TMEM131L/MALT1/CD70/IL23A/TNFSF11/CD86/HAVCR2/ABL1/TMIGD2/TNFRSF1B/IL18/CSK/ZMIZ1/CCR2/ERBB2/VNN1/NLRP3/IHH/HHLA2/PLA2G2D/DPP4/CD47/PELI1/CD274/TNFSF13B/CASP3/FOXN1/IFNG/LAT/FANCA/PRDM1/IL23R/LGALS9B/IL7/HLA-G/FAM49B/ZBTB7B/ADORA2A/PYCARD/EFNB1/CD4/HLA-DMB/TNFAIP8L2/LGALS9/IL12A/IL2/FYN/SMAD7/CD80/SELENOK/IL2RA/GLI3/EGR3/IL10/AP3D1/AP3B1/DUSP22/ZBTB1/ZC3H8/CGAS/CD160/HMGB1/EFNB2/CLPTM1/LILRB4/RAC1/SH3RF1/HLA-E/PTPN2/XCL1/VCAM1/CYLD  
## GO:0051249 FCRL3/CTLA4/CD19/BTLA/CAMK4/IRF4/MZB1/TIGIT/LAX1/SIRPG/KLRK1/PTPRC/CLECL1/CARD11/PTPN22/IKZF3/RIPOR2/CD6/INPP5D/CD28/TNFRSF13B/RUNX3/SLA2/RASAL3/LCK/GRAP2/PDCD1/RASGRP1/CCDC88B/ZAP70/CD3E/TNFRSF13C/TESPA1/SIT1/IGLL5/DOCK8/MYB/IL7R/CD2/CORO1A/TBX21/SASH3/CD5/RAC2/HLA-DPB1/SPN/IL12RB1/LAG3/BTK/CD40LG/LILRB1/TNFSF8/ICOS/VAV1/FANCD2/TNFAIP3/NOD2/HLA-F/FOXP3/HLA-DOA/BTN2A2/ATM/BANK1/TOX/CD27/TNFRSF18/JAK3/FCGR2B/ITPKB/NCKAP1L/PTPN6/SAMSN1/IL4R/CD300A/CR1/CCR7/HLA-DPA1/SIRPB1/TNFSF4/CD83/CD74/IGLL1/NFATC2/PAG1/ADAM8/TMEM131L/NFAM1/MALT1/CD22/CD70/GPR183/IL23A/TNFSF11/CD86/HAVCR2/ABL1/TMIGD2/TNFRSF1B/IL18/ATAD5/CSK/ZMIZ1/CCR2/ERBB2/VNN1/IL27RA/HMGB3/NLRP3/CD40/IHH/HHLA2/PLA2G2D/DPP4/CD47/PELI1/CD274/TNFSF13B/CASP3/FOXN1/CD38/IFNG/LAT/FANCA/PRDM1/IL23R/LGALS9B/IL7/SLAMF8/HLA-G/SLC39A10/FAM49B/ZBTB7B/ADORA2A/PYCARD/LST1/TLR9/PGLYRP2/EFNB1/CLEC7A/CD4/HLA-DMB/TNFAIP8L2/LGALS9/IL12A/RHBDD3/IL2/FYN/SMAD7/CD80/SELENOK/ATP11C/TAC1/IL2RA/GLI3/EGR3/IL10/AP3D1/AP3B1/DUSP22/ZBTB1/MPL/TYROBP/ZC3H8/CGAS/CD160/HMGB1/EFNB2/CLPTM1/LILRB4/PPP2R3C/RAC1/SH3RF1/HLA-E/FLT3LG/PTPN2/XCL1/VCAM1/SHLD2/CYLD/MLH1  
## GO:0007159 CTLA4/BTLA/TIGIT/LAX1/SIRPG/KLRK1/PTPRC/CLECL1/CARD11/PTPN22/RIPOR2/CD6/CD28/RUNX3/SELL/RASAL3/LCK/GRAP2/PDCD1/RASGRP1/CCDC88B/ZAP70/ITGA4/ITGAL/CD3E/TNFRSF13C/TESPA1/DOCK8/MYB/IL7R/CORO1A/TBX21/SASH3/CD5/RAC2/HLA-DPB1/SPN/IL12RB1/LAG3/CD40LG/LILRB1/ICOS/VAV1/ADD2/NOD2/FOXP3/BTN2A2/CD27/JAK3/FCGR2B/SEMA4D/ITPKB/NCKAP1L/PTPN6/IL4R/CD300A/CCR7/HLA-DPA1/SIRPB1/TNFSF4/ITGB7/CD83/CD74/SKAP1/PAG1/ADAM8/TMEM131L/SELPLG/STK10/ADTRP/MALT1/CD70/IL23A/TNFSF11/CD86/HAVCR2/TMIGD2/IL18/CSK/ZMIZ1/CCR2/ERBB2/VNN1/NLRP3/FERMT3/IHH/HHLA2/ITGB2/PLA2G2D/DPP4/CD47/PELI1/CD274/TNFSF13B/CASP3/IFNG/IL23R/LGALS9B/TNF/IL7/HLA-G/FAM49B/ZBTB7B/ADORA2A/PYCARD/EFNB1/CD4/HLA-DMB/TNFAIP8L2/LGALS9/IL12A/IL2/FYN/SMAD7/CD80/GCNT1/SELENOK/ETS1/IL2RA/GLI3/EGR3/IL10/AP3D1/AP3B1/DUSP22/ZBTB1/ZC3H8/CD160/HMGB1/EFNB2/PTAFR/LILRB4/RAC1/MADCAM1/HLA-E/PTPN2/XCL1/VCAM1/CYLD/ITGB1  
## GO:0022407 CTLA4/BTLA/TIGIT/LAX1/SIRPG/KLRK1/PTPRC/CLECL1/CARD11/PTPN22/RIPOR2/CD6/CD28/RUNX3/RASAL3/LCK/GRAP2/PDCD1/RASGRP1/CCDC88B/ZAP70/ITGA4/CD3E/TNFRSF13C/TESPA1/AKNA/DOCK8/MYB/IL7R/CORO1A/TBX21/SASH3/CD5/HLA-DPB1/SPN/IL12RB1/LAG3/CD40LG/LILRB1/ICOS/VAV1/NOD2/FOXP3/BTN2A2/CXCL13/CD27/JAK3/FCGR2B/WNT1/ITPKB/NCKAP1L/PTPN6/IL4R/CD300A/CCR7/TNR/HLA-DPA1/SIRPB1/TNFSF4/CD83/CD74/SKAP1/PAG1/ADAM8/TMEM131L/ADTRP/MALT1/CD70/IL23A/TNFSF11/CD86/HAVCR2/ABL1/TMIGD2/IL18/CSK/ZMIZ1/CCR2/ERBB2/ADAM19/VNN1/FXYD5/NLRP3/FERMT3/IHH/HHLA2/ITGB2/PLA2G2D/DPP4/CD47/MDGA1/PELI1/CD274/TNFSF13B/CASP3/IFNG/IL23R/LGALS9B/TNF/IL7/HLA-G/JAK2/ALOX15/FAM49B/VEGFA/ZBTB7B/ADORA2A/PYCARD/EFNB1/CD4/HLA-DMB/TNFAIP8L2/MYO10/LGALS9/IL12A/IL2/PTK2/FYN/SMAD7/CD80/LEF1/SELENOK/DMTN/MAPK7/ETS1/IL2RA/NF2/GLI3/EGR3/IL10/AP3D1/AP3B1/DUSP22/CELSR2/ZBTB1/ZC3H8/BMP6/KIF26B/CD160/FLOT1/HMGB1/EFNB2/EPCAM/LRFN3/PTAFR/LILRB4/TENM3/RAC1/SPINT2/HLA-E/PTPN2/XCL1/VCAM1/CDH1/CYLD/PRKCA  
## Count  
## GO:0042110 196  
## GO:0030098 153  
## GO:0050863 141  
## GO:0051249 180  
## GO:0007159 140  
## GO:0022407 156

## ##################################  
## Comparison: SFIvsELI   
## ONTOLOGY ID Description GeneRatio  
## GO:0042110 BP GO:0042110 T cell activation 146/2047  
## GO:0030098 BP GO:0030098 lymphocyte differentiation 116/2047  
## GO:0050863 BP GO:0050863 regulation of T cell activation 105/2047  
## GO:0030217 BP GO:0030217 T cell differentiation 85/2047  
## GO:0007159 BP GO:0007159 leukocyte cell-cell adhesion 103/2047  
## GO:0051249 BP GO:0051249 regulation of lymphocyte activation 126/2047  
## BgRatio pvalue p.adjust qvalue  
## GO:0042110 464/18670 8.211639e-34 4.940943e-30 4.142988e-30  
## GO:0030098 353/18670 6.401722e-29 1.925958e-25 1.614919e-25  
## GO:0050863 314/18670 5.630087e-27 1.129208e-23 9.468424e-24  
## GO:0030217 240/18670 5.665029e-24 8.521620e-21 7.145391e-21  
## GO:0007159 337/18670 4.910919e-23 5.909799e-20 4.955375e-20  
## GO:0051249 485/18670 5.466524e-21 5.188821e-18 4.350834e-18  
## geneID  
## GO:0042110 RIPOR2/NLRC3/BCL11B/FANCD2/DOCK8/SLAMF6/TCF7/MYB/CTLA4/BTLA/CARD11/ITK/CCR6/PTPN6/TMEM131L/RASAL3/TESPA1/BTN2A2/GRAP2/LILRB1/CAMK4/CLECL1/SIRPG/KLRK1/ITPKB/CCDC88B/BTN3A1/IL7R/LAX1/DOCK2/JAML/TNFSF8/ZAP70/CORO1A/CASP8/CCR7/CD40LG/WAS/LCK/SIT1/CD300A/CD6/GPR18/SPN/PTPN22/PTPRC/TIGIT/CD2/PIK3CD/SASH3/CD3G/ZMIZ1/ICOS/ITGAL/IL12RB1/SIRPB1/THEMIS/LAT/TREML2/CD3E/F2RL1/PIK3CG/PDCD1/RHOH/JAK3/RASGRP1/FUT7/CD28/CD3D/PAG1/CSK/VAV1/ABL1/RUNX3/APBB1IP/TBX21/WNT1/IL6/NCKAP1L/CD27/LCP1/IRF4/HLA-DPB1/PYCARD/AIRE/P2RX7/IL23A/HLA-DOA/FANCA/TMIGD2/TNFSF4/FOXP3/RARA/SLA2/FCGR2B/RUNX2/VNN1/CCR2/NOD2/CR1/ZBTB1/ADAM8/TNFRSF13C/RAC2/CCR9/EOMES/SH3RF1/LEF1/CASP3/ERBB2/IL18/GATA3/LY9/CD83/NFATC2/CD5/TNFSF11/SDC4/NLRP3/LAG3/TNFAIP8L2/MALT1/FOXN1/TCIRG1/BATF/HAVCR2/SRC/CD86/STAT3/PRKAR1A/GLI3/CD8A/IL4R/CD7/FYN/IL4/HLA-DMB/CD4/HLA-DPA1/ADORA2A/TNFRSF1B/LIG4/PRDM1/CD8B/ELF4/TNFRSF14  
## GO:0030098 HDAC9/BCL11B/FANCD2/SLAMF6/TCF7/MYB/CTLA4/CARD11/ITK/INPP5D/CCR6/PLCG2/PTPN6/PTK2B/TMEM131L/TESPA1/BTN2A2/ATM/POU2F2/CD79B/IKZF1/CAMK4/DOCK10/AICDA/ITPKB/IL7R/DOCK2/TNFSF8/ZAP70/CCR7/FCRL3/CD40LG/LCK/MS4A1/GPR18/SPN/BTK/PTPN22/PTPRC/ITGA4/CD2/IKZF3/PIK3CD/SASH3/CD3G/ZMIZ1/IL12RB1/TOX/THEMIS/CD3E/CD19/RHOH/JAK3/RASGRP1/FUT7/CD28/CD3D/VAV1/ABL1/RUNX3/TBX21/WNT1/IL6/NCKAP1L/BLNK/CD27/NFAM1/IRF4/AIRE/IL23A/HLA-DOA/FANCA/C17orf99/DCLRE1C/TNFSF4/FOXP3/RARA/FCGR2B/RUNX2/VNN1/CCR2/CR1/ZBTB1/ADAM8/CCR9/EOMES/SH3RF1/LEF1/ERBB2/IL18/GATA3/LYL1/ITGB1/LY9/CD83/NFATC2/NLRP3/LAG3/MALT1/FOXN1/TCIRG1/CR2/BATF/CD86/STAT3/GLI3/CD8A/IL4R/IL4/SLAMF8/CD4/FLT3/PGLYRP2/ADGRG3/LIG4/PRDM1  
## GO:0050863 RIPOR2/FANCD2/DOCK8/MYB/CTLA4/BTLA/CARD11/PTPN6/TMEM131L/RASAL3/TESPA1/BTN2A2/GRAP2/LILRB1/CAMK4/CLECL1/SIRPG/KLRK1/ITPKB/CCDC88B/IL7R/LAX1/TNFSF8/ZAP70/CORO1A/CCR7/CD40LG/LCK/SIT1/CD300A/CD6/SPN/PTPN22/PTPRC/TIGIT/CD2/SASH3/ZMIZ1/ICOS/IL12RB1/SIRPB1/LAT/CD3E/PDCD1/JAK3/RASGRP1/CD28/PAG1/CSK/VAV1/ABL1/RUNX3/TBX21/IL6/NCKAP1L/CD27/IRF4/HLA-DPB1/PYCARD/IL23A/HLA-DOA/FANCA/TMIGD2/TNFSF4/FOXP3/RARA/FCGR2B/VNN1/CCR2/NOD2/CR1/ZBTB1/ADAM8/TNFRSF13C/RAC2/SH3RF1/CASP3/ERBB2/IL18/GATA3/CD83/NFATC2/CD5/TNFSF11/SDC4/NLRP3/LAG3/TNFAIP8L2/MALT1/FOXN1/HAVCR2/SRC/CD86/PRKAR1A/GLI3/IL4R/FYN/IL4/HLA-DMB/CD4/HLA-DPA1/ADORA2A/TNFRSF1B/PRDM1/TNFRSF14  
## GO:0030217 BCL11B/FANCD2/SLAMF6/TCF7/MYB/CTLA4/CARD11/ITK/CCR6/TMEM131L/TESPA1/BTN2A2/CAMK4/ITPKB/IL7R/DOCK2/TNFSF8/ZAP70/CCR7/LCK/GPR18/SPN/PTPN22/PTPRC/CD2/PIK3CD/SASH3/CD3G/ZMIZ1/IL12RB1/THEMIS/CD3E/RHOH/JAK3/RASGRP1/FUT7/CD28/CD3D/VAV1/ABL1/RUNX3/TBX21/WNT1/IL6/NCKAP1L/CD27/IRF4/AIRE/IL23A/HLA-DOA/FANCA/TNFSF4/FOXP3/RARA/RUNX2/VNN1/CCR2/CR1/ZBTB1/ADAM8/CCR9/EOMES/SH3RF1/LEF1/ERBB2/IL18/GATA3/LY9/CD83/NFATC2/NLRP3/LAG3/MALT1/FOXN1/TCIRG1/BATF/CD86/STAT3/GLI3/CD8A/IL4R/IL4/CD4/LIG4/PRDM1  
## GO:0007159 RIPOR2/DOCK8/MYB/CTLA4/BTLA/CARD11/SELL/PTPN6/TMEM131L/RASAL3/TESPA1/BTN2A2/GRAP2/LILRB1/CLECL1/SIRPG/KLRK1/ITPKB/CCDC88B/IL7R/LAX1/ZAP70/CORO1A/CCR7/CD40LG/LCK/CD300A/CD6/SPN/PTPN22/SEMA4D/PTPRC/ITGA4/TIGIT/SASH3/ZMIZ1/ICOS/ITGAL/ADD2/IL12RB1/SIRPB1/STK10/CD3E/PDCD1/JAK3/RASGRP1/CD28/PAG1/CSK/VAV1/RUNX3/TBX21/IL6/SKAP1/NCKAP1L/ADTRP/CD27/HLA-DPB1/PYCARD/ITGB7/IL23A/TMIGD2/TNFSF4/FOXP3/RARA/FCGR2B/VNN1/CCR2/NOD2/ZBTB1/ADAM8/TNFRSF13C/TNF/RAC2/SELPLG/CASP3/ERBB2/IL18/GATA3/ITGB1/CD83/CD5/TNFSF11/SDC4/NLRP3/LAG3/TNFAIP8L2/MALT1/HAVCR2/SRC/CD86/PRKAR1A/GLI3/IL4R/FYN/IL4/HLA-DMB/CD4/HLA-DPA1/ADORA2A/ITGB2/TNFRSF14/FERMT3  
## GO:0051249 RIPOR2/FANCD2/DOCK8/MYB/CTLA4/BTLA/CARD11/INPP5D/BANK1/PTPN6/TMEM131L/RASAL3/TESPA1/BTN2A2/ATM/GRAP2/LILRB1/CAMK4/CLECL1/SIRPG/CD22/KLRK1/ITPKB/CCDC88B/IL7R/LAX1/TNFSF8/ZAP70/CORO1A/CCR7/FCRL3/CD40LG/LCK/SIT1/CD300A/CD6/SPN/BTK/PTPN22/PTPRC/HLA-F/TIGIT/CD2/IKZF3/SASH3/ZMIZ1/ICOS/IL12RB1/SIRPB1/TOX/LAT/CD3E/PDCD1/CD19/JAK3/RASGRP1/IL27RA/CD28/PAG1/CSK/VAV1/ABL1/RUNX3/TBX21/IL6/NCKAP1L/CD27/NFAM1/ATAD5/IRF4/HLA-DPB1/PYCARD/IL23A/HLA-DOA/FANCA/TMIGD2/TNFSF4/FOXP3/RARA/SLA2/TNFAIP3/FCGR2B/VNN1/CCR2/NOD2/CR1/ZBTB1/ADAM8/TNFRSF13C/RAC2/SH3RF1/CASP3/ERBB2/IL18/GATA3/CD83/NFATC2/CD5/TNFSF11/SDC4/NLRP3/LAG3/TNFAIP8L2/MALT1/FOXN1/LST1/SAMSN1/HAVCR2/SRC/CD86/PRKAR1A/GLI3/IL4R/FYN/IL4/SLAMF8/HLA-DMB/CD4/HLA-DPA1/ADORA2A/TNFRSF1B/CD40/PGLYRP2/PRDM1/PKN1/TNFRSF14  
## Count  
## GO:0042110 146  
## GO:0030098 116  
## GO:0050863 105  
## GO:0030217 85  
## GO:0007159 103  
## GO:0051249 126

BiocManager::install("clusterProfiler")

## Bioconductor version 3.11 (BiocManager 1.30.10), R 4.0.0 (2020-04-24)

## Installing package(s) 'clusterProfiler'

## Warning: package 'clusterProfiler' is in use and will not be installed

## Installation path not writeable, unable to update packages: boot, class,  
## foreign, KernSmooth, MASS, nlme, nnet, spatial, survival

## Old packages: 'limma', 'vctrs'